(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 January 2001 (04.01.2001)

PCT

(10) International Publication Number
WO 01/00673 A1

(51) International Patent Classification[7]: C07K 14/47, C07H 21/04, C12N 15/63, 1/2, C12P 21/02

(21) International Application Number: PCT/US00/18198

(22) International Filing Date: 29 June 2000 (29.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/345,464 30 June 1999 (30.06.1999) US

(71) Applicant: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors: BARNES, Thomas, M.; 22 Hanson Street #2, Boston, MA 02118 (US). FRASER, Christopher, C.; 52 Grassland Street, Lexington, MA 02421 (US). WRIGHTON, Nicholas; 18 Lloyd Street, Winchester, MA 01890 (US). MYERS, Paul; 14 Cornelius Way, Cambridge, MA 02141 (US). BUSFIELD, Samantha, J.; Apartment 1, 15 Trowbridge Street, Cambridge, MA 02138 (US). SHARP, John, D.; 245 Park Avenue, Arlington, MA 02476 (US).

(74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/00673 A1

(54) Title: MEMBRANE-ASSOCIATED AND SECRETED PROTEINS AND USES THEREOF

(57) Abstract: The invention provides isolated nucleic acid molecules, designated INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378 which encode wholly secreted or membrane-associated proteins. The invention also provides antisense nucleic acid molecules, expression vectors containing the nucleic acid molecules of the invention, host cells into which the expression vectors have been introduced, and non-human transgenic animals in which a nucleic acid molecule of the invention has been introduced or disrupted. The invention still further provides isolated polypeptides, fusion polypeptides, antigenic peptides and antibodies. Diagnostic, screening and therapeutic methods utilizing compositions of the invention are also provided.

MEMBRANE-ASSOCIATED AND SECRETED PROTEINS AND USES THEREOF

This application claims priority to co-pending U.S. Application No. 09/345,464, filed June 30, 1999, the entire contents of which are incorporated herein by reference in its entirety.

Background of the Invention

Many secreted proteins, for example, cytokines, play a vital role in the regulation of cell growth, cell differentiation, and a variety of specific cellular responses. A number of medically useful proteins, including erythropoietin, granulocyte-macrophage colony stimulating factor, human growth hormone, and various interleukins, are secreted proteins.

Many membrane-associated proteins are receptors which bind a ligand and transduce an intracellular signal, leading to a variety of cellular responses. The identification and characterization of such a receptor enables one to identify both the ligands which bind to the receptor and the intracellular molecules and signal transduction pathways associated with the receptor, permitting one to identify or design modulators of receptor activity, *e.g.*, receptor agonists or antagonists and modulators of signal transduction.

Thus, an important goal in the design and development of new therapies is the identification and characterization of membrane-associated and secreted proteins and the genes which encode them.

Summary of the Invention

The present invention is based, at least in part, on the discovery of cDNA molecules encoding INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378 all of which are either wholly secreted or transmembrane proteins. These proteins, fragments, derivatives, and variants thereof are collectively referred to as "polypeptides of the invention" or "proteins of the invention." Nucleic acid molecules encoding the polypeptides or proteins of the invention are collectively referred to as "nucleic acids of the invention."

The nucleic acids and polypeptides of the present invention are useful as modulating agents in regulating a variety of cellular processes. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding a polypeptide of the invention or a biologically active portion thereof. The present invention also provides nucleic acid molecules which are suitable for use as primers or hybridization probes for the detection of nucleic acids encoding a polypeptide of the invention.

The invention features nucleic acid molecules which are at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number 207178 (the "cDNA of ATCC® Accession Number 207178"), the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number PTA-249 (the "cDNA of ATCC® Accession Number PTA-249"), or the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number PTA-250 (the "cDNA of ATCC® Accession Number PTA-250"), or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2400, 2600, 2800, 3000, 3200, 3400, 3600, 3800, or 4000) nucleotides of the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, the nucleotide sequence of the cDNA of ATCC® Accession Number 207178, the nucleotide sequence of the cDNA of ATCC® Accession Number PTA-249, or the nucleotide sequence of the cDNA of ATCC® Accession Number PTA-250, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, the amino acid sequence encoded by the cDNA of ATCC® Accession Number 207178, the amino acid sequence encoded by the cDNA of ATCC® Accession Number PTA-249, or the amino acid sequence encoded by the cDNA of ATCC® Accession Number PTA-250.

In preferred embodiments, the nucleic acid molecules have the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, the nucleotide sequence of the cDNA of ATCC® Accession Number 207178, the nucleotide sequence of the cDNA of ATCC® Accession Number PTA-249, or the nucleotide sequence of the cDNA of ATCC® Accession Number PTA-250, or a complement thereof.

Also within the invention are nucleic acid molecules which encode a fragment of a polypeptide having the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, or a fragment including at least 15 (25, 30, 50, 100, 150, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, or 1400) contiguous amino acids of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, the amino acid sequence encoded by the cDNA of ATCC® Accession Number 207178, the amino acid sequence encoded by the cDNA of ATCC® Accession Number PTA-249, or the amino acid sequence encoded by the cDNA of ATCC® Accession Number PTA-250.

The invention includes nucleic acid molecules which encode a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, the amino acid sequence encoded by the cDNA of ATCC® Accession Number 207178, the amino acid sequence encoded by the cDNA of ATCC® Accession Number PTA-249, or the amino acid sequence encoded by the cDNA of ATCC® Accession Number PTA-250, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule consisting of a nucleic acid sequence encoding SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, the nucleotide sequence of the cDNA of ATCC® Accession Number 207178, the nucleotide sequence of the cDNA of ATCC® Accession Number PTA-249, or the nucleotide sequence of the cDNA of ATCC® Accession Number PTA-250, or a complement thereof under stringent conditions.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 60%, preferably 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, the amino acid sequence encoded by the cDNA of ATCC® Accession Number 207178, the amino acid sequence encoded by the cDNA of ATCC® Accession Number PTA-249, or the amino acid sequence encoded by the cDNA of ATCC® Accession Number PTA-250.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 60%, preferably 65%, 75%, 85%, or 95% identical the nucleic acid sequence encoding SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, or complement thereof, the non-coding strand of the cDNA of ATCC® Accession Number 207178, the non-coding strand of the cDNA of ATCC® Accession Number PTA-249, or the non-coding strand of the cDNA of ATCC® Accession Number PTA-250.

Also within the invention are polypeptides which are naturally occurring allelic variants of a polypeptide that includes the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, the amino acid sequence encoded by the cDNA of ATCC® Accession Number 207178, the amino acid sequence encoded by the cDNA of ATCC® Accession Number PTA-249, or the amino acid sequence encoded by the cDNA of ATCC® Accession Number PTA-250, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule having the sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, or a complement thereof, under stringent conditions. Such allelic variant differ at 1%, 2%, 3%, 4%, or 5% of the amino acid residues.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, the cDNA of ATCC® Accession Number 207178, the cDNA of ATCC® Accession Number PTA-249, or the cDNA of ATCC® Accession Number PTA-250, or a complement thereof. In other embodiments, the nucleic acid molecules are at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000, or 4200) nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule consisting of the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, the cDNA of ATCC® Accession Number 207178, the cDNA of ATCC® Accession Number PTA-249, or the cDNA of ATCC® Accession Number PTA-250, or a complement thereof.

In other embodiments, the isolated nucleic acid molecules encode an extracellular, transmembrane, or cytoplasmic domain of a polypeptide of the invention.

In another embodiment, the invention provides an isolated nucleic acid molecule which is antisense to the coding strand of a nucleic acid of the invention.

Another aspect of the invention provides vectors, *e.g.*, recombinant expression vectors, comprising a nucleic acid molecule of the invention. In another embodiment, the invention provides host cells containing such a vector or a nucleic acid molecule of the invention. The invention also provides methods for producing a polypeptide of the invention by culturing, in a suitable medium, a host cell of the invention containing a recombinant expression vector such that a polypeptide is produced.

Another aspect of this invention features isolated or recombinant proteins and polypeptides of the invention. Preferred proteins and polypeptides possess at least one biological activity possessed by the corresponding naturally-occurring human polypeptide. An activity, a biological activity, or a functional activity of a polypeptide or nucleic acid of the invention refers to an activity exerted by a protein, polypeptide or nucleic acid molecule of the invention on a responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques. Such activities can be a direct activity, such as an association with or an enzymatic activity on a second protein or an indirect activity, such as a cellular signaling activity mediated by interaction of the protein with a second protein.

In one embodiment, the isolated polypeptide of the invention lacks both a transmembrane and a cytoplasmic domain. In another embodiment, the polypeptide lacks both a transmembrane domain and a cytoplasmic domain and is soluble under physiological conditions.

For INTERCEPT 340, biological activities include, *e.g.*, (1) the ability to form protein-protein interactions with proteins in the signaling pathway of the naturally-

occurring polypeptide; (2) the ability to bind a ligand of the naturally-occurring polypeptide; (3) the ability to interact with an INTERCEPT 340 receptor, *e.g.*, a cell surface receptor (*e.g.*, an integrin); (4) the ability to modulate the activity of an intracellular molecule that participates in a signal transduction pathway, *e.g.*, an intracellular molecule in the integrin signalling (*e.g.*, a cdk2 inhibitor); (5) the ability to assemble into fibrils; (6) the ability to strengthen and organize the extracellular matrix; (7) the ability to modulate the shape of tissues and cells; (8) the ability to interact with (*e.g.*, bind to) components of the extracellular matrix; and (9) the ability to modulate cell migration. Other activities include the ability to modulate function, survival, morphology, migration, proliferation and/or differentiation of cells of tissues in which it is expressed (*e.g.*, splenic cells). For example, additional biological activities of INTERCEPT 340 include: (1) the ability to modulate splenic cell activity; (2) the ability to modulate skeletal morphogenesis; and/or (3) the ability to modulate smooth muscle cell proliferation and differentiation.

For MANGO 003, biological activities include, *e.g.*, (1) the ability to form protein-protein (*e.g.*, protein-ligand) interactions with proteins in the signaling pathway of the naturally-occurring polypeptide; (2) the ability to interact with (*e.g.*, bind to) a ligand of the naturally-occurring polypeptide; (3) the ability to interact with a MANGO 003 receptor, *e.g.*, a cell surface receptor; (4) the ability to modulate cell surface recognition; (5) the ability to transduce an extracellular signal (*e.g.*, by interacting with a ligand and/or a cell-surface receptor); (6) the ability to modulate a signal transduction pathway; and (7) the ability to modulate signal transmission at a chemical synapse. Other activities include the ability to modulate function, survival, morphology, proliferation and/or differentiation of cells of tissues in which it is expressed (*e.g.*, thyroid, liver, skeletal muscle, kidney, heart, lung, testis and brain). For example, the activities of MANGO 003 can include modulation of endocrine, hepatic, skeletal muscular, renal, cardiovascular, reproductive and/or brain function.

For MANGO 347, biological activities include, *e.g.*, (1) the ability to form protein-protein interactions with proteins in the signaling pathway of the naturally-occurring polypeptide; (2) the ability to interact with a ligand of the naturally-occurring polypeptide; (3) the ability to interact with a MANGO 347 receptor; and (4) the ability to modulate a developmental process, *e.g.*, morphogenesis, cellular migration, adhesion, proliferation, differentiation, and/or survival. Other activities include the ability to modulate function, survival, morphology, proliferation and/or differentiation of cells of tissues in which it is expressed (*e.g.*, brain cells). For example, the activities of MANGO 347 can include modulation of neural (*e.g.*, CNS) function.

For TANGO 272, biological activities include, *e.g.*, (1) the ability to form protein-protein interactions with proteins in the signaling pathway of the naturally-occurring

polypeptide; (2) the ability to bind a ligand of the naturally-occurring polypeptide; (3) the ability to interact with a TANGO 272 receptor, *e.g.*, a cell surface receptor (*e.g.*, an integrin); (4) the ability to modulate cell-cell contact; (5) the ability to modulate cell attachment; (6) the ability to modulate cell fate; and (7) the ability to modulate tissue repair and/or wound healing. Other activities include the ability to modulate function, survival, morphology, proliferation and/or differentiation of cells of tissues in which it is expressed (*e.g.*, microvascular endothelial cells). For example, the activities of MANGO 347 can include modulation of cardiovascular function.

For TANGO 295, biological activities include, *e.g.*, (1) the ability to form protein-protein interactions with proteins in the signaling pathway of the naturally-occurring polypeptide; (2) the ability to bind a ligand of the naturally-occurring polypeptide; (3) the ability to interact with a TANGO 295 receptor; (4) the ability to interact with (*e.g.*, bind to) a nucleic acid; and (5) the ability to elicit pyrimidine-specific endonuclease activity. Other activities include the ability to modulate function, survival, morphology, proliferation and/or differentiation of cells of tissues in which it is expressed (*e.g.*, mammary epithelium).

For TANGO 354, biological activities include, *e.g.*, (1) the ability to form protein-protein interactions with proteins in the signaling pathway of the naturally-occurring polypeptide; (2) the ability to bind a ligand of the naturally-occurring polypeptide; (3) the ability to interact with (*e.g.*, bind to) a TANGO 354 receptor, *e.g.*, a cell surface receptor; (4) the ability to modulate cell surface recognition; (5) the ability to modulate cellular motility, *e.g.*, chemotaxis and/or chemokinesis; (6) the ability to transduce an extracellular signal (*e.g.*, by interacting with a ligand and/or a cell-surface receptor); and (7) the ability to modulate a signal transduction pathway. Other activities include the ability to modulate function, survival, morphology, proliferation and/or differentiation of cells of tissues in which it is expressed (*e.g.*, hematopoietic tissues). For example, TANGO 354 biological activities can further include: (1) regulation of hematopoiesis; (2) modulation (*e.g.*, increasing or decreasing) of haemostasis; (3) modulation of an inflammatory response; (4) modulation of neoplastic growth, *e.g.*, inhibition of tumor growth; and (5) modulation of thrombolysis.

For TANGO 378, biological activities include, *e.g.*, (1) the ability to form protein-protein interactions with proteins in the signaling pathway of the naturally-occurring polypeptide; (2) the ability to bind a ligand of the naturally-occurring polypeptide; (3) the ability to interact with a TANGO 378 receptor; (4) the ability to transduce an extracellular signal; and (5) the ability to modulate a signal transduction pathway (*e.g.*, adenylate cyclase, or phosphatidylinositol 4,5-bisphosphate (PIP_2), inositol 1,4,5-triphosphate (IP_3)). Other activities include the ability to modulate function, survival, morphology, proliferation

and/or differentiation of cells of tissues in which it is expressed (*e.g.*, natural killer cells). For example, TANGO 378 biological activities can further include the ability to modulate an immune response in a subject, for example, (1) by modulating immune cytotoxic responses against pathogenic organisms, *e.g.*, viruses, bacteria, and parasites; (2) by modulating organ rejection after transplantation; and (3) by modulating immune recognition and lysis of normal and malignant cells.

In one embodiment, a polypeptide of the invention has an amino acid sequence sufficiently identical to an identified domain of a polypeptide of the invention. As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (*e.g.*, with a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have a common structural domain and/or common functional activity. For example, amino acid or nucleotide sequences which contain a common structural domain having about 60% identity, preferably 65% identity, more preferably 75%, 85%, 95%, 98% or more identity are defined herein as sufficiently identical.

In one embodiment, a MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, or TANGO 378 polypeptide of the invention includes a signal peptide.

In another embodiment, a nucleic acid molecule of the invention encodes a MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, or TANGO 378 polypeptide which includes a signal peptide.

In another embodiment, a MANGO 003, TANGO 272, TANGO 354, or TANGO 378 polypeptide of the invention includes one or more of the following domains: (1) a signal peptide; (2) an N-terminal extracellular domain; (3) a C-terminal transmembrane domain; and (4) a cytoplasmic domain.

The polypeptides of the present invention, or biologically active portions thereof, can be operably linked to a heterologous amino acid sequence to form fusion proteins. In one embodiment, the fusion protein consists of a chimeric protein assembled from portions of the protein from different species.

In one embodiment, the isolated polypeptide of the invention lacks both a transmembrane and a cytoplasmic domain. In another embodiment, the polypeptide lacks both a transmembrane domain and a cytoplasmic domain and is soluble under physiological conditions.

The invention further features antibodies that specifically bind a polypeptide of the invention such as monoclonal or polyclonal antibodies. In addition, the polypeptides of the invention or biologically active portions thereof, or antibodies of the invention, can be

incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides methods for detecting the presence of the activity or expression of a polypeptide of the invention in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of activity such that the presence of activity is detected in the biological sample.

In another aspect, the invention provides methods for modulating activity of a polypeptide of the invention comprising contacting a cell with an agent that modulates (inhibits or stimulates) the activity or expression of a polypeptide of the invention such that activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to a polypeptide of the invention.

In another embodiment, the agent modulates expression of a polypeptide of the invention by modulating transcription, splicing, or translation of an mRNA encoding a polypeptide of the invention. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of an mRNA encoding a polypeptide of the invention.

The present invention also provides methods to treat a subject having a disorder characterized by aberrant activity of a polypeptide of the invention or aberrant expression of a nucleic acid of the invention by administering an agent which is a modulator of the activity of a polypeptide of the invention or a modulator of the expression of a nucleic acid of the invention to the subject. In one embodiment, the modulator is a protein of the invention. In another embodiment, the modulator is a nucleic acid of the invention. In other embodiments, the modulator is a peptide, peptidomimetic, or other small organic molecule. The present invention also provides diagnostic assays for identifying the presence or absence of a genetic lesion or mutation characterized by at least one of: (i) aberrant modification or mutation of a gene encoding a polypeptide of the invention, (ii) mis-regulation of a gene encoding a polypeptide of the invention, and (iii) aberrant post-translational modification of the invention wherein a wild-type form of the gene encodes a protein having the activity of the polypeptide of the invention.

In another aspect, the invention provides a method for identifying a compound that binds to or modulates the activity of a polypeptide of the invention. In general, such methods entail measuring a biological activity of the polypeptide in the presence and absence of a test compound and identifying those compounds which alter the activity of the polypeptide.

The invention also features methods for identifying a compound which modulates the expression of a polypeptide or nucleic acid of the invention by measuring the expression of the polypeptide or nucleic acid in the presence and absence of the compound.

In yet a further aspect, the invention provides substantially purified antibodies or fragments thereof including human and non-human antibodies or fragments thereof which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29 or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207178, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-249, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-250; a fragment of at least 15 amino acid residues of the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29; an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. In various embodiments, the substantially purified antibodies of the invention, or fragments thereof can be human, non-human, chimeric and/or humanized antibodies.

Any of the antibodies of the invention can be conjugated to a therapeutic moiety or to a detectable substance. Non-limiting examples of detectable substances that can be conjugated to the antibodies of the invention are an enzyme, a prosthetic group, a fluorescent material, a luminescent material, a bioluminescent material, and a radioactive material.

The invention also provides a kit containing an antibody of the invention conjugated to a detectable substance, and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody of the invention and a pharmaceutically acceptable carrier. In preferred embodiments, the pharmaceutical composition contains an antibody of the invention, a therapeutic moiety, and a pharmaceutically acceptable carrier.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figures 1A-1B depict the cDNA sequence of human INTERCEPT 340 (SEQ ID NO:1) and the predicted amino acid sequence of INTERCEPT 340 (SEQ ID NO:2). The

open reading frame of SEQ ID NO:1 extends from nucleotide 1222 to nucleotide 1944 of SEQ ID NO:1 (SEQ ID NO:3).

Figure 2 depicts a hydropathy plot of human INTERCEPT 340. Relatively hydrophobic residues are above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace. Below the hydropathy plot, the numbers corresponding to the amino acid sequence of INTERCEPT 340 are indicated. The amino acid sequence of each of the fibrillar collagen C-terminal domains are indicated by underlining and the abbreviation "COLF".

Figure 3 depicts an alignment of each of the fibrillar collagen C-terminal domains (also referred to herein as "COLF domains") of human INTERCEPT 340 with consensus hidden Markov model COLF domains. For each alignment, the upper sequence is the consensus amino acid sequence (SEQ ID NOs:31, 32, and 33), while the lower sequence amino acid sequence corresponds to amino acid 58 to amino acid 116 of SEQ ID NO:2 (SEQ ID NO:34), amino acid 126 to amino acid 151 of SEQ ID NO:2 (SEQ ID NO:35), and amino acid 186 to amino acid 217 of SEQ ID NO:2 (SEQ ID NO:36).

Figures 4A-4C depict the cDNA sequence of human MANGO 003 (SEQ ID NO:4) and the predicted amino acid sequence of MANGO 003 (SEQ ID NO:5). The open reading frame of SEQ ID NO:4 extends from nucleotide 57 to nucleotide 1568 of SEQ ID NO:4 (SEQ ID NO:6).

Figure 5 depicts a hydropathy plot of human MANGO 003. Relatively hydrophobic residues are above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace. Below the hydropathy plot, the numbers corresponding to the amino acid sequence of MANGO 003 are indicated. The amino acid sequence of each of the immunoglobulin domains, and the neurotransmitter gated ion channel domain are indicated by underlining and the abbreviations "ig" and "neur chan", respectively.

Figure 6 depicts an alignment of each of the immunoglobulin domains (also referred to herein as "Ig domains") of human MANGO 003 with the consensus hidden Markov model immunoglobulin domains. For each alignment, the upper sequence is the consensus sequence (SEQ ID NO:37), while the lower sequence corresponds to amino acid 44 to amino acid 101 of SEQ ID NO:5 (SEQ ID NO:38), amino acid 165 to amino acid 223 of SEQ ID NO:5 (SEQ ID NO:39), and amino acid 261 to amino acid 340 of SEQ ID NO:5 (SEQ ID NO:40).

Figure 7 depicts an alignment of the neurotransmitter gated ion channel domain of human MANGO 003 with the consensus hidden Markov model neurotransmitter gated ion

channel domain. The upper sequence is the consensus sequence (SEQ ID NO:42), while the lower sequence corresponds to amino acid 388 amino acid 397 of SEQ ID NO:5 (SEQ ID NO:43).

Figure 8 depicts the cDNA sequence of mouse MANGO 003 (SEQ ID NO:7) and the predicted amino acid sequence of MANGO 003 (SEQ ID NO:8). The open reading frame of SEQ ID NO:7 extends from nucleotide 1 to nucleotide 626 of SEQ ID NO:4 (SEQ ID NO:9).

Figure 9 depicts a hydropathy plot of mouse MANGO 003. Relatively hydrophobic residues are above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace. Below the hydropathy plot, the numbers corresponding to the amino acid sequence of mouse MANGO 003 are indicated.

Figure 10 depicts the cDNA sequence of human MANGO 347 (SEQ ID NO:10) and the predicted amino acid sequence of MANGO 347 (SEQ ID NO:11). The open reading frame of SEQ ID NO:10 extends from nucleotide 31 to nucleotide 444 of SEQ ID NO:10 (SEQ ID NO:12).

Figure 11 depicts a hydropathy plot of human MANGO 347. Relatively hydrophobic residues are above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. Below the hydropathy plot, the numbers corresponding to the amino acid sequence of MANGO 347 are indicated. The amino acid sequence of the CUB domain is indicated by underlining and the abbreviation "CUB".

Figure 12 depicts an alignment of the CUB domain of human MANGO 347 with a consensus hidden Markov model CUB domain. The upper sequence is the consensus amino acid sequence (SEQ ID NO:44), while the lower sequence corresponds to amino acid 40 to amino acid 136 of SEQ ID NO:11 (SEQ ID NO:45).

Figures 13A-13D depict the cDNA sequence of human TANGO 272 (SEQ ID NO:13) and the predicted amino acid sequence of TANGO 272 (SEQ ID NO:14). The open reading frame of SEQ ID NO:13 extends from nucleotide 230 to nucleotide 3379 of SEQ ID NO:13 (SEQ ID NO:15).

Figure 14 depicts a hydropathy plot of human TANGO 272. Relatively hydrophobic residues are above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace. Below the hydropathy plot, the numbers corresponding to the amino acid sequence of

TANGO 272 are indicated. The amino acid sequence of each of the fourteen EGF-like domains and the delta serrate ligand domain is indicated by underlining and the abbreviation "EGF-like" and "DSL", respectively.

Figures 15A-15C depict an alignment of each of the EGF-like domains of human TANGO 272 with consensus hidden Markov model EGF-like domains. The upper sequence is the consensus amino acid sequence (SEQ ID NO:46), while the lower sequence corresponds to amino acid 151 to amino acid 181 of SEQ ID NO:14 (SEQ ID NO:49); amino acid 200 to amino acid 229 of SEQ ID NO:14 (SEQ ID NO:50); amino acid 242 to amino acid 272 of SEQ ID NO:14 (SEQ ID NO:51); amino acid 285 to amino acid 315 of SEQ ID NO:14 (SEQ ID NO:52); amino acid 328 to amino acid 358 of SEQ ID NO:14 (SEQ ID NO:53); amino acid 378 to amino acid 404 of SEQ ID NO:14 (SEQ ID NO:54); amino acid 417 to amino acid 447 of SEQ ID NO:14 (SEQ ID NO:55); amino acid 460 to amino acid 490 of SEQ ID NO:14 (SEQ ID NO:56); amino acid 503 to amino acid 533 of SEQ ID NO:14 (SEQ ID NO:57); amino acid 546 to amino acid 576 of SEQ ID NO:14 (SEQ ID NO:58); amino acid 589 to amino acid 619 of SEQ ID NO:14 (SEQ ID NO:59); amino acid 632 to amino acid 661 of SEQ ID NO:14 (SEQ ID NO:60); amino acid 674 to amino acid 704 of SEQ ID NO:14 (SEQ ID NO:61); and amino acid 717 amino acid 747 of SEQ ID NO:14 (SEQ ID NO:62). For alignment of the delta serrate ligand domain, the upper sequence is the consensus hidden Markov model (SEQ ID NO:47), while the lower sequence corresponds to amino acid 518 to amino acid 576 of SEQ ID NO:14 (SEQ ID NO:63).

Figures 16A-16B depict the cDNA sequence of mouse TANGO 272 (SEQ ID NO:16) and the predicted amino acid sequence of TANGO 272 (SEQ ID NO:17). The open reading frame of SEQ ID NO:16 extends from nucleotide 1 to nucleotide 1492 of SEQ ID NO:16 (SEQ ID NO:18).

Figure 17 depicts a hydropathy plot of mouse TANGO 272. Relatively hydrophobic residues are above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace. Below the hydropathy plot, the numbers corresponding to the amino acid sequence of mouse TANGO 272 are indicated.

Figure 18 depicts the cDNA sequence of human TANGO 295 (SEQ ID NO:22) and the predicted amino acid sequence of TANGO 295 (SEQ ID NO:23). The open reading frame of SEQ ID NO:22 extends from nucleotide 217 to nucleotide 684 of SEQ ID NO:28 (SEQ ID NO:24).

Figure 19 depicts a hydropathy plot of human TANGO 295. Relatively hydrophobic residues are above the dashed horizontal line, and relatively hydrophilic

residues are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace. Below the hydropathy plot, the numbers corresponding to the amino acid sequence of human TANGO 295 are indicated. The amino acid sequence of the pancreatic ribonuclease domain is indicated by underlining and the abbreviation "RNase A".

Figure 20 depicts an alignment of the pancreatic ribonuclease domain of human TANGO 295 with a consensus hidden Markov model pancreatic ribonuclease domain. The upper sequence is the consensus amino acid sequence (SEQ ID NO:96), while the lower sequence corresponds to amino acid 32 to amino acid 156 of SEQ ID NO:23 (SEQ ID NO:97).

Figures 21A-21B depict the cDNA sequence of human TANGO 354 (SEQ ID NO:25) and the predicted amino acid sequence of TANGO 354 (SEQ ID NO:26). The open reading frame of SEQ ID NO:25 extends from nucleotide 62 to nucleotide 976 of SEQ ID NO:25 (SEQ ID NO:27).

Figure 22 depicts a hydropathy plot of human TANGO 354. Relatively hydrophobic residues are above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace. Below the hydropathy plot, the numbers corresponding to the amino acid sequence of human TANGO 354 are indicated. The amino acid sequence of the immunoglobulin domain is indicated by underlining and the abbreviation "ig".

Figure 23 depicts an alignment of the immunoglobulin domain of human TANGO 354 with a consensus hidden Markov model immunoglobulin domains. The upper sequence is the consensus amino acid sequence (SEQ ID NO:37), while the lower sequence corresponds to amino acid 33 to amino acid 110 of SEQ ID NO:26 (SEQ ID NO:41).

Figures 24A-24C depict the cDNA sequence of human TANGO 378 (SEQ ID NO:28) and the predicted amino acid sequence of TANGO 378 (SEQ ID NO:29). The open reading frame of SEQ ID NO:28 extends from nucleotide 42 to nucleotide 1625 of SEQ ID NO:28 (SEQ ID NO:30).

Figure 25 depicts a hydropathy plot of human TANGO 378. Relatively hydrophobic residues are above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace. Below the hydropathy plot, the numbers corresponding to the amino acid sequence of human TANGO 378 are indicated. The amino acid sequence of the seven transmembrane domain is indicated by underlining and the abbreviation "7tm".

Figure 26 depicts an alignment of the seven transmembrane receptor domain of human TANGO 378 with a consensus hidden Markov model of this domain. The upper sequence is the consensus amino acid sequence (SEQ ID NO:98), while the lower sequence corresponds to amino acid 187 to amino acid 515 of SEQ ID NO:29 (SEQ ID NO:99).

Figures 27A-27C depict a global alignment between the nucleotide sequence of the open reading frame (ORF) of human MANGO 003 (SEQ ID NO:6) and the nucleotide sequence of the open reading frame of mouse MANGO 003 (SEQ ID NO:9). The upper sequence is the human MANGO 003 ORF nucleotide sequence, while the lower sequence is the mouse MANGO 003 ORF nucleotide sequence. These nucleotides sequences share a 31.1% identity. The global alignment was performed using the ALIGN program version 2.0u (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of –1212; Myers and Miller, 1989, *CABIOS* 4:11-7).

Figures 28A-28B depict a local alignment between the nucleotide sequence of human MANGO 003 (SEQ ID NO:4) and the nucleotide sequence of mouse MANGO 003 (SEQ ID NO:7). The upper sequence is the human MANGO 003 nucleotide sequence, while the lower sequence is the mouse MANGO 003 nucleotide sequence. These nucleotides sequences share a 62.8 % identity over nucleotide 970 to nucleotide 2080 of the human MANGO 003 sequence (nucleotide 10 to nucleotide 1070 of mouse MANGO 003). The local alignment was performed using the L-ALIGN program version 2.0u54 July 1996 (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a score of 3241; Huang and Miller, 1991, *Adv. Appl. Math.* 12:373-381).

Figure 29 depicts a global alignment between the amino acid sequence of human MANGO 003 (SEQ ID NO:5) and the amino acid sequence of mouse MANGO 003 (SEQ ID NO:8). The upper sequence is the human MANGO 003 amino acid sequence, while the lower sequence is the mouse MANGO 003 amino acid sequence. These amino acid sequences share a 30.1% identity. The global alignment was performed using the ALIGN program version 2.0u (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of –488; Myers and Miller, 1989, *CABIOS* 4:11-7).

Figures 30A-30E depict a global alignment between the nucleotide sequence of the open reading frame (ORF) of human TANGO 272 (SEQ ID NO:15) and the nucleotide sequence of the open reading frame of mouse TANGO 272 (SEQ ID NO:18). The upper sequence is the mouse TANGO 272 ORF nucleotide sequence, while the lower sequence is the human TANGO 272 ORF nucleotide sequence. These nucleotides sequences share a 39.1% identity. The global alignment was performed using the ALIGN program version 2.0u (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of –79; Myers and Miller, 1989, *CABIOS* 4:11-7).

Figures 31A-31D depict a local alignment between the nucleotide sequence of human TANGO 272 (SEQ ID NO:13) and the nucleotide sequence of mouse TANGO 272 (SEQ ID NO:16). The upper sequence is the human TANGO 272 nucleotide sequence, while the lower sequence is the mouse TANGO 272 nucleotide sequence. These nucleotides sequences share a 67.6 % identity over nucleotide 1890 to nucleotide 4610 of the human TANGO 272 sequence (nucleotide 10 to nucleotide 2560 of mouse TANGO 272). The local alignment was performed using the L-ALIGN program version 2.0u54 July 1996 (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a score of 8462; Huang and Miller, 1991, *Adv. Appl. Math.* 12:373-381).

Figures 32A-32B depict a global alignment between the amino acid sequence of human TANGO 272 (SEQ ID NO:14) and the amino acid sequence of mouse TANGO 272 (SEQ ID NO:17). The upper sequence is the human TANGO 272 amino acid sequence, while the lower sequence is the mouse TANGO 272 amino acid sequence. These amino acid sequences share a 38.2% identity. The global alignment was performed using the ALIGN program version 2.0u (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of –19; Myers and Miller, 1989, *CABIOS* 4:11-7).

Figures 33A-33D depict the cDNA sequence of rat TANGO 272 (SEQ ID NO:19) and the predicted amino acid sequence of TANGO 272 (SEQ ID NO:20). The open reading frame of SEQ ID NO:19 extends from nucleotide 925 to nucleotide 2832 of SEQ ID NO:19 (SEQ ID NO:21).

Figures 34A-34H depict a global alignment between the nucleotide sequence of human TANGO 272 (SEQ ID NO:13) and the nucleotide sequence of rat TANGO 272 (SEQ ID NO:19). The upper sequence is the human TANGO 272 nucleotide sequence, while the lower sequence is the rat TANGO 272 nucleotide sequence. These nucleotides sequences share a 55.7% identity. The global alignment was performed using the ALIGN program version 2.0u (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of 8635; Myers and Miller, 1989, *CABIOS* 4:11-7).

Figures 35A-35F depict a global alignment between the nucleotide sequence of mouse TANGO 272 (SEQ ID NO:16) and the nucleotide sequence of rat TANGO 272 (SEQ ID NO:19). The upper sequence is the mouse TANGO 272 nucleotide sequence, while the lower sequence is the rat TANGO 272 nucleotide sequence. These nucleotides sequences share a 43.7% identity. The global alignment was performed using the ALIGN program version 2.0u (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of 2827; Myers and Miller, 1989, *CABIOS* 4:11-7).

Figure 36 depicts a global alignment of the human TANGO 295 and GenPept AF037081 amino acid sequences. The upper sequence is the human TANGO 295 sequence (SEQ ID NO:23), while the lower sequence is the GenPept AF037081 sequence (SEQ ID

NO:100). GenPept AF037081 encodes a ribonuclease k6 protein. The global alignment revealed a 53.2% identity between these two sequences (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of 405; Myers and Miller, 1989, *CABIOS* 4:11-7).

Figures 37A-37C depict a global alignment of the human TANGO 295 (SEQ ID NO:22) and GenPept AF037081 (SEQ ID NO:100) nucleotide sequences. The upper sequence is the human TANGO 295 sequence, while the lower sequence is the GenPept AF037081 sequence. The global alignment revealed a 22.6% identity between these two sequences (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of -2718; Myers and Miller, 1989, *CABIOS* 4:11-7).

Figures 38A-38B depict a local alignment of the human TANGO 295 (SEQ ID NO:22) and GenPept AF037081 (SEQ ID NO:100) nucleotide sequences. The upper sequence is the human TANGO 295 sequence, while the lower sequence is the GenPept AF037081 sequence. The local alignment revealed a 62.7% identity between nucleotide 235 to nucleotide 687 of human TANGO 295, and nucleotide 3 to nucleotide 453 of AF037081; 43.4% identity between nucleotide 410 to nucleotide 850 of human TANGO 295, and nucleotide 3 to nucleotide 450 of AF037081; and 46.5% identity between nucleotide 432 to nucleotide 700 of human TANGO 295, and nucleotide 5 to nucleotide 251 of AF037081 (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of 1214; Huang and Miller, 1991, *Adv. Appl. Math.* 12:373-381).

Figures 39A-39B depict an alignment of each of the EGF-like domains and laminin-EGF-like domains of mouse TANGO 272 with consensus hidden Markov model EGF-like domains. For alignments of the EGF-like domains, the upper sequence is the consensus amino acid sequence (SEQ ID NO:46), while the lower sequence corresponds to amino acids 37-67 of SEQ ID NO:17 (SEQ ID NO:64); amino acid 80 to amino acid 110 of SEQ ID NO:17 (SEQ ID NO:65); amino acid 123 to amino acid 153 of SEQ ID NO:17 (SEQ ID NO:66); and amino acid 166 to amino acid 196 of SEQ ID NO:17 (SEQ ID NO:67). For alignments of the laminin/EGF-like domains, the upper sequence is the consensus hidden Markov model domain (SEQ ID NO:48), while the lower sequence corresponds to amino acid 3 to amino acid 37 of SEQ ID NO:17 (SEQ ID NO:68); amino acid 41 to amino acid 80 of SEQ ID NO:17 (SEQ ID NO:69); amino acid 83 to amino acid 123 of SEQ ID NO:17 (SEQ ID NO:70); and amino acid 127 to amino acid 172 of SEQ ID NO:17 (SEQ ID NO:71). For alignment of the delta serrate ligand domain, the upper sequence is the consensus hidden Markov model domain (SEQ ID NO:47), while the lower sequence corresponds to amino acid 10 to amino acid 67 of SEQ ID NO:17 (SEQ ID NO:72).

Figure 40 depicts a hydropathy plot of rat TANGO 272. Relatively hydrophobic residues are above the dashed horizontal line, and relatively hydrophilic residues are below

the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace. Below the hydropathy plot, the numbers corresponding to the amino acid sequence of rat TANGO 272 are indicated.

Figures 41A-41D depict an alignment of each of the EGF-like domains and laminin-EGF-like domains of rat TANGO 272 with consensus hidden Markov model of EGF-like domains. For alignments of the EGF-like domains, the upper sequence is the consensus amino acid sequence (SEQ ID NO:46), while the lower sequence corresponds to amino acid 18 to amino acid 48 of SEQ ID NO:20 (SEQ ID NO:73); amino acid 61 to amino acid 91 of SEQ ID NO:20 (SEQ ID NO:74); amino acids 105-137 of SEQ ID NO:20 (SEQ ID NO:75); amino acids 150-180 of SEQ ID NO:20 (SEQ ID NO:76); amino acids 193-223 of SEQ ID NO:20 (SEQ ID NO:77); amino acids 236-266 of SEQ ID NO:20 (SEQ ID NO:78); amino acids 279-309 of SEQ ID NO:20 (SEQ ID NO:79); amino acids 322-352 of SEQ ID NO:20 (SEQ ID NO:80); amino acids 365-394 of SEQ ID NO:20 (SEQ ID NO:81); amino acids 407-437 of SEQ ID NO:20 (SEQ ID NO:82); and amino acids 450-480 of SEQ ID NO:20 (SEQ ID NO:83). For alignments of the laminin/EGF-like domains, the upper sequence is the consensus hidden Markov model domain (SEQ ID NO:48), while the lower sequence corresponds to amino acids 22-61 of SEQ ID NO:20 (SEQ ID NO:84); amino acids 65-105 of SEQ ID NO:20 (SEQ ID NO:85); amino acids 109-150 of SEQ ID NO:20 (SEQ ID NO:86); amino acids 154-193 of SEQ ID NO:20 (SEQ ID NO:87); amino acids 197-236 of SEQ ID NO:20 (SEQ ID NO:88); amino acids 240-279 of SEQ ID NO:20 (SEQ ID NO:89); amino acids 283-322 of SEQ ID NO:20 (SEQ ID NO:90); amino acids 326-365 of SEQ ID NO:20 (SEQ ID NO:91); amino acids 368-407 of SEQ ID NO:20 (SEQ ID NO:92); amino acids 411-450 of SEQ ID NO:20 (SEQ ID NO:93); and amino acids 454-489 of SEQ ID NO:20 (SEQ ID NO:94). For alignment of the delta serrate ligand domain, the upper sequence is the consensus hidden Markov model domain (SEQ ID NO:47), while the lower sequence corresponds to amino acids 246-309 of SEQ ID NO:20 (SEQ ID NO:95).

Detailed Description of the Invention

The present invention is based, at least in part, on the discovery of cDNA molecules encoding INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378, all of which are either wholly secreted or transmembrane proteins.

The proteins and nucleic acid molecules of the present invention comprise a family of molecules having certain conserved structural and functional features. As used herein, the term "family" is intended to mean two or more proteins or nucleic acid molecules

having a common structural domain and having sufficient amino acid or nucleotide sequence identity as defined herein. Family members can be from either the same or different species. For example, a family can comprise two or more proteins of human origin, or can comprise one or more proteins of human origin and one or more of non-human origin. Members of the same family may also have common structural domains.

For example, INTERCEPT 340 family members can include at least one, preferably two, and more preferably three fibrillar collagen C-terminal domains (also referred to herein as "COLF domains"). As used herein, a "fibrillar collagen C-terminal domain" refers to an amino acid sequence of about 15 to 65, preferably about 20-60, more preferably about 25, 31-58 amino acids in length. Consensus hidden Markov model COLF domains contain the sequence of SEQ ID NOs:31, 32, and 33 (Figure 3). The more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues in the consensus sequence are indicated by lowercase letters. A comparison of the C-terminal sequences of fibrillar collagens, collagens X, VIII, and the collagen C1q revealed a conserved cluster of amino acid residues having aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine) that exhibited marked similarities in hydrophilicity profiles between the different collagens, despite a low level of sequence similarity. These similarities in hydrophilicity profiles within their C-termini suggest that these proteins may adopt a common tertiary structure and that the conserved cluster of aromatic residues in this domain may be involved in C-terminal trimerization. The COLF domains of INTERCEPT 340 extend from about amino acids 58 to 116, 126 to 151, and 186 to 217 of SEQ ID NO:2 (SEQ ID NOs:34, 35, and 36, respectively) (Figure 3). By alignment of the amino acid sequence of the consensus hidden Markov model COLF amino acid sequence with the amino acid sequence of the COLF domains of INTERCEPT 340, conserved amino acid residues having aromatic side chains can be found. For example, conserved tyrosine, tryptophan and phenylalanine residues can be found at amino acid 87, 88 and 133 of SEQ ID NO:2.

MANGO 003 and TANGO 354 family members can include at least one, preferably two, and more preferably three immunoglobulin domains. As used herein, an "immunoglobulin domain" (also referred to herein as "Ig") refers to an amino acid sequence of about 45 to 85, preferably about 55-80, more preferably about 57, 58, or 78, 79 amino acids in length. Preferably, the immunoglobulin domains have a bit score for the alignment of the sequence to the Ig family Hidden Markov Model (HMM) of at least 10, preferably 20-30, more preferably 22-40, more preferably 40-50, 50-75, 75-100, 100-200 or greater. The Ig family HMM has been assigned the PFAM Accession PF00047. Consensus hidden Markov model immunoglobulin domains are shown Figures 6 and 23 (SEQ ID NO:37). The more conserved residues in the consensus sequence are indicated by uppercase letters

and the less conserved residues in the consensus sequence are indicated by lowercase letters. Immunoglobulin domains are present in a variety of proteins (including secreted and membrane-associated proteins). Membrane-associated proteins may be involved in protein-protein, and protein-ligand interaction at the cell surface, and thus may influence diverse activities including cell surface recognition and/or signal transduction. The immunoglobulin domains of MANGO 003 extend from about amino acids 44 to 101, 165 to 223, and 261 to 240 of SEQ ID NO:5 (SEQ ID NOs:38, 39, and 40, respectively) (Figure 6). The immunoglobulin domain of TANGO 354 extend from about amino acids 33 to 110 of SEQ ID NO:26 (SEQ ID NO:41) (Figure 23).

MANGO 003 family member can include a neurotransmitter-gated ion channel domain. As used herein, a "neurotransmitter-gated ion channel domain" refers to an amino acid sequence of about 5 to 20, preferably about 7 to 12, more preferably about 9 to 10 amino acids in length. The neurotransmitter-gated ion channel domain HMM has been assigned the PFAM Accession PF00065. A consensus hidden Markov model neurotransmitter-gated ion channel domain contain the sequence of SEQ ID NO:42 shown in Figure 7. The more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues in the consensus sequence are indicated by lowercase letters. The neurotransmitter-gated ion channel domains of MANGO 003 extend from about amino acids 388 to 397 of SEQ ID NO:5 (SEQ ID NO:43).

TANGO 272 family members can include at least one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, preferably thirteen, and more preferably fourteen EGF-like domains. Preferably, the EGF-like domains are found in the extracellular domain of a TANGO 272 protein. As used herein, an "EGF-like domain" refers to an amino acid sequence of about 25 to 50, preferably about 30 to 45, and more preferably 30 to 40 amino acid residues in length. An EGF domain further contains at least about 2 to 10, preferably, 3 to 9, 4 to 8, or 6 to 7 conserved cysteine residues. A consensus hidden Markov model EGF-like domain sequence includes six cysteines, all of which are thought to be involved in disulfide bonds having the following amino acid sequence: Cys-Xaa(5, 7)-Cys-Xaa(4, 5, 12)-Cys-Xaa(1, 5, 6)-Cys-Xaa(1)-Cys-Xaa(1)- Cys-Xaa(8)-Cys (SEQ ID NO:46), where Xaa is any amino acid. The region between the fifth and the sixth cysteine typically contains two conserved glycines of which at least one is present in most EGF-like domains.

In one embodiment, TANGO 272 includes at least one EGF-like domain having the sequences selected from the group consisting of: amino acids 151-181 of SEQ ID NO:14 (SEQ ID NO:49); amino acids 200-229 of SEQ ID NO:14 (SEQ ID NO:50); amino acids 242-272 of SEQ ID NO:14 (SEQ ID NO:51); amino acids 285-315 of SEQ ID NO:14 (SEQ ID NO:52); amino acids 328-358 of SEQ ID NO:14 (SEQ ID NO:53); amino acids 378-404 of SEQ ID NO:14 (SEQ ID NO:54); amino acids 417-447 of SEQ ID NO:14 (SEQ ID

NO:55); amino acids 460-490 of SEQ ID NO:14 (SEQ ID NO:56); amino acids 503-533 of SEQ ID NO:14 (SEQ ID NO:57); amino acids 546-576 of SEQ ID NO:14 (SEQ ID NO:58); amino acids 589-619 of SEQ ID NO:14 (SEQ ID NO:59); amino acids 632-661 of SEQ ID NO:14 (SEQ ID NO:60); amino acids 674-704 of SEQ ID NO:14 (SEQ ID NO:61); and amino acids 717-747 of SEQ ID NO:14 (SEQ ID NO:62).

In another embodiment, TANGO 272 includes at least one EGF-like domain having the sequences selected from the group consisting of: 37-67 of SEQ ID NO:17 (SEQ ID NO:64); amino acids 80-110 of SEQ ID NO:17 (SEQ ID NO:65); amino acids 123-153 of SEQ ID NO:17 (SEQ ID NO:66); and amino acids 166-196 of SEQ ID NO:17 (SEQ ID NO:67).

In yet another embodiment, TANGO 272 includes at least one EGF-like domain having the sequences selected from the group consisting of: amino acids 18-48 of SEQ ID NO:20 (SEQ ID NO:73); amino acids 61-91 of SEQ ID NO:20 (SEQ ID NO:74); amino acids 105-137 of SEQ ID NO:20 (SEQ ID NO:75); amino acids 150-180 of SEQ ID NO:20 (SEQ ID NO:76); amino acids 193-223 of SEQ ID NO:20 (SEQ ID NO:77); amino acids 236-266 of SEQ ID NO:20 (SEQ ID NO:78); amino acids 279-309 of SEQ ID NO:20 (SEQ ID NO:79); amino acids 322-352 of SEQ ID NO:20 (SEQ ID NO:80); amino acids 365-394 of SEQ ID NO:20 (SEQ ID NO:81); amino acids 407-437 of SEQ ID NO:20 (SEQ ID NO:82); and amino acids 450-480 of SEQ ID NO:20 (SEQ ID NO:83).

An alignment of the consensus hidden Markov model EGF-like domains with the EGF-like domains of human TANGO 272 is shown in Figures 15A-15C. The more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues in the consensus sequence are indicated by lowercase letters. By alignment of the amino acid sequence of the consensus hidden Markov model EGF-like domain with the amino acid sequence of the EGF-like domains of TANGO 272, conserved cysteine residues can be found. For example, conserved cysteine residues can be found at amino acid 151, 159, 164, 167, 200, 206, 211, 218, 220, 229, 242, 249, 263, 264, 272, 285, 291, 297, 304, 306, 315, 328, 334, 340, 347, 349, 358, 378, 386, 393, 395, 404, 417, 423, 429, 436, 438, 447, 460, 466, 472, 479, 481, 490, 503, 509, 515, 522, 524, 533, 546, 552, 558, 565, 567, 576, 589, 595, 601, 608, 610, 619, 632, 637, 643, 650, 652, 661, 674, 680, 686, 693, 695, 717, 723, 729, 736, 738 and 747 of SEQ ID NO:14.

TANGO 272 family members can include at least one delta serrate ligand domain. As used herein, a "delta serrate ligand domain" (also referred to herein as a "DSL domain") refers to an amino acid sequence of about 30-70, more preferably 45-60, and most preferably 58 amino acids in length typically found in transmembrane signaling molecules that regulate differentiation in metazoans (Lissemore et al., 1999, *Mol. Phylogenet. Evol.* 11(2):308-19). In one embodiment, human TANGO 272 includes a delta serrate ligand

domain from about amino acids 518 to 576 of SEQ ID NO:14 (SEQ ID NO:63); and about amino acids 246 to 309 of SEQ ID NO:20 (SEQ ID NO:95). Figure 15B depicts an alignment of the consensus hidden Markov model delta serrate ligand domain (SEQ ID NO:47) with this domain in human TANGO 272 at amino acids 518 to 576 of SEQ ID NO:14 (SEQ ID NO:63). Figures 39A-39B depict an alignment of the consensus hidden Markov model delta serrate ligand domain (SEQ ID NO:47) with this domain in mouse TANGO 272 at amino acids 10 to 67 of SEQ ID NO:17 (SEQ ID NO:72). Figures 41A-41B depict an alignment of the consensus hidden Markov model delta serrate ligand domain (SEQ ID NO:47) with this domain in rat TANGO 272 at amino acids 246 to 309 of SEQ ID NO:20 (SEQ ID NO:95).

TANGO 272 family members can include at least one RGD cell attachment site. As used herein, the term "RGD cell attachment site" refers to a cell adhesion sequence consisting of amino acids Arg-Gly-Asp typically found in extracellular matrix proteins such as collagens, laminin and fibronectin, among others (reviewed in Ruoslahti, 1996, *Annu. Rev. Cell Dev. Biol.* 12:697-715). Preferably, the RGD cell attachment site is located in the extracellular domain of a TANGO 272 protein and interacts (*e.g.*, binds to) a cell surface receptor, such as an integrin receptor. As used herein, the term "integrin" refers to a family of receptors comprising α/β heterodimers that mediate cell attachment to extracellular matrices and cell-cell adhesion events. The α subunits vary in size between 120 and 180 kDa and are each noncovalently associated with a β subunit (90-110 kDa) (reviewed by Hynes, 1992, *Cell* 69:11-25). Most integrins are expressed in a wide variety of cells, and most cells express several integrins. There are at least 8 known α subunits and 14 known β subunits. The majority of the integrin ligands are extracellular matrix proteins involved in substratum cell adhesion such as collagens, laminin, fibronectin among others. The RGD cell attachment site is located at about amino acid residues 177-179 of SEQ ID NO:14.

MANGO 347 family members can include a CUB domain sequence. As used herein, the term "CUB domain" includes an amino acid sequence having at least about 80-150, preferably 90-130, more preferably 96-120, and most preferably about 110 amino acids in length. Preferably, a CUB domain further includes at least one, preferably two, three, and most preferably four conserved cysteine residues. Preferably, the conserved cysteine residues form at least one, and preferably two disulfide bridges (*e.g.*, Cys1-Cys2, and Cys3-Cys4) resulting in a β-barrel configuration. The CUB domain of MANGO 347 extends from about amino acid 40 to amino acid 136 of SEQ ID NO:11 (SEQ ID NO:45). Figure 12 depicts an alignment of the consensus hidden Markov model CUB domain (SEQ ID NO:44) with this domain in human MANGO 347 at amino acids 40 to 136 of SEQ ID NO:11 (SEQ ID NO:45).

TANGO 295 family members can include a pancreatic ribonuclease domain sequence. As used herein, the term "pancreatic ribonuclease domain" includes an amino acid sequence having at least about 100 to 150, preferably 110-140, more preferably 120-130, and most preferably 124 amino acids in length. Preferably, a pancreatic ribonuclease domain further includes at least one, preferably two, three, four and most preferably five conserved cysteine residues and an amino acid residue, *e.g.*, a lysine, which is involved in catalytic activity. Preferably, at least one cysteine residue is involved in a disulfide bond, a lysine residue is involved in catalytic activity, and three other residues involved in substrate binding. Proteins having the pancreatic ribonuclease domain are pyrimidine-specific endonucleases present in high quantities in the pancreas of a number of mammalian taxa and of a few reptiles. The pancreatic ribonuclease domain of TANGO 295 extends from about amino acid 32 to amino acid 156 of SEQ ID NO:23 (SEQ ID NO:97). Figure 20 depicts an alignment of the consensus hidden Markov model pancreatic ribonuclease domain (SEQ ID NO:96) with this domain in human TANGO 295 at amino acids 32 to 156 of SEQ ID NO:23 (SEQ ID NO:97).

Based on structural similarities, TANGO 378 family members can be classified as members of the superfamily of G-protein coupled receptor. As used herein, the term "G protein-coupled receptor" or "GPCR" refers to a family of proteins that preferably comprise an N-terminal extracellular domain, seven transmembrane domains (also referred to as membrane-spanning domains), three extracellular domains (also referred to as extracellular loops), three cytoplasmic domains (also referred to as cytoplasmic loops), and a C-terminal cytoplasmic domain (also referred to as a cytoplasmic tail). Members of the GPCR family also share certain conserved amino acid residues, some of which have been determined to be critical to receptor function and/or G protein signaling. An alignment of the transmembrane domains of 44 representative GPCRs can be found at http://mgdkk1.nidll.nih.gov:8000/extended.html.

Accordingly, in one embodiment, TANGO 378 family members can include at least one, two, three, four, five, six, or preferably, seven transmembrane domains, and thus has a "7 transmembrane receptor profile". As used herein, the term "7 transmembrane receptor profile" includes an amino acid sequence having at least about 10-300, preferably about 15-200, more preferably about 20-100 amino acid residues, or at least about 22-100 amino acids in length and having a bit score for the alignment of the sequence to the 7tm_1 family Hidden Markov Model (HMM) of at least 10, preferably 20-30, more preferably 22-40, more preferably 40-50, 50-75, 75-100, 100-200 or greater. The 7tm_1 family HMM has been assigned the PFAM Accession PF00001 (http://genome.wustl.edu/Pfam/WWWdata/7tm_1.html). In one embodiment, the seven transmembrane domains of TANGO 378 extend from about amino acids 245 to about

amino acid 269 of SEQ ID NO:29 (SEQ ID NO:135), about amino acids 287 to about amino acid 306 of SEQ ID NO:29 (SEQ ID NO:136), about amino acids 323 to about amino acid 343 of SEQ ID NO:29 (SEQ ID NO:137), about amino acids 358 to about amino acid 376 of SEQ ID NO:29 (SEQ ID NO:138), about amino acids 414 to about amino acid 438 of SEQ ID NO:29 (SEQ ID NO:139), about amino acids 457 to about amino acid 477 of SEQ ID NO:29 (SEQ ID NO:140), and about amino acids 485 to about amino acid 504 of SEQ ID NO:29 (SEQ ID NO:141); and a C-terminal cytoplasmic domain which extends from about amino acid 505 to amino acid 528 of SEQ ID NO:29 (SEQ ID NO:142). Figure 26 depicts an alignment of each of the transmembrane domains of TANGO 378 with the consensus hidden Markov model seven transmembrane receptor domain (SEQ ID NO:98).

To identify the presence of a 7 transmembrane receptor profile in a TANGO 378, the amino acid sequence of the protein is searched against a database of HMMs (*e.g.*, the Pfam database, release 2.1) using the default parameters (http://www.sanger.ac.uk/Software/Pfam/HMM_search). For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program for PF00001 and score of 15 is the default threshold score for determining a hit. Alternatively, the seven transmembrane domain can be predicted based on stretches of hydrophobic amino acids forming α-helices (SOUSI server). Accordingly, proteins having at least 50-60% identity, preferably about 60-70%, more preferably about 70-80%, or about 80-90% identity with the 7 transmembrane receptor profile of human TANGO 378 are within the scope of the invention.

TANGO 378 family members can include at least one, preferably two, and most preferably three extracellular loops. As defined herein, the term "loop" includes an amino acid sequence having a length of at least about 4, preferably about 5-10, preferably about 10-20, and more preferably about 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, or 100-150 amino acid residues, and has an amino acid sequence that connects two transmembrane domains within a protein or polypeptide. Accordingly, the N-terminal amino acid of a loop is adjacent to a C-terminal amino acid of a transmembrane domain in a naturally-occurring TANGO 378 or TANGO 378-like molecule, and the C-terminal amino acid of a loop is adjacent to an N-terminal amino acid of a transmembrane domain in a naturally-occurring TANGO 378 or TANGO 378-like molecule. As used herein, an "extracellular loop" includes an amino acid sequence located outside of a cell, or extracellularly. For example, an extracellular loop can be found at about amino acids 307-322, 377-413, and 478-484 of SEQ ID NO:29.

TANGO 378 family members can include at least one, preferably two, and most preferably three cytoplasmic loops. As used herein, a "cytoplasmic loop" includes an amino

acid sequence located within a cell or within the cytoplasm of a cell. For example, a cytoplasmic loop is found at about amino acids 270-286, 344-357, and 439-456 of SEQ ID NO:29.

In one embodiment, a MANGO 003, a TANGO 272, a TANGO 354 or a TANGO 378 family member can include one or more of the following domains: (1) an N-terminal extracellular domain, (2) a transmembrane domain, or (3) a C-terminal cytoplasmic domain.

MANGO 003, a TANGO 272, a TANGO 354 or a TANGO 378 family member can include an extracellular domain. When located at the N-terminal domain the extracellular domain is referred to herein as an "N-terminal extracellular domain" or an "extracellular domain". As used herein, an "N-terminal extracellular domain" includes an amino acid sequence having about 1-800, preferably about 1-746, more preferably about 1-650, more preferably about 1-550, more preferably about 1-369, about 150 amino acid residues in length and is located outside of a cell or extracellularly. The C-terminal amino acid residue of a "N-terminal extracellular domain" is adjacent to an N-terminal amino acid residue of a transmembrane domain in a naturally-occurring MANGO 003, TANGO 272, TANGO 354 or TANGO 378 protein. Preferably, the N-terminal extracellular domain is capable of interacting (*e.g.*, binding to) with an extracellular signal, for example, a ligand (*e.g.*, a glycoprotein hormone) or a cell surface receptor (*e.g.*, an integrin receptor). Most preferably, the N-terminal extracellular domain mediates a variety of biological processes, for example, protein-protein interactions, signal transduction and/or cell adhesion. In one embodiment, an N-terminal cytoplasmic domain is located at about amino acids 25-374 of SEQ ID NO:5 (SEQ ID NO:103); about amino acids 1-73 of SEQ ID NO:8 (SEQ ID NO:107); at about amino acids 21-767 of SEQ ID NO:14 (SEQ ID NO:114); at about amino acids 1-216 of SEQ ID NO:17 (SEQ ID NO:118); at about amino acids 1-500 of SEQ ID NO:20 (SEQ ID NO:122); at about amino acids 20-169 of SEQ ID NO:26 (SEQ ID NO:129); and at about amino acids 22-244 of SEQ ID NO:29 (SEQ ID NO:134).

In another embodiment, a MANGO 003, a TANGO 272, a TANGO 354 or a TANGO 378 family member can include a transmembrane domain. As used herein, the term "transmembrane domain" includes an amino acid sequence of about 15 amino acid residues in length which spans the plasma membrane. More preferably, a transmembrane domain includes about at least 20, 25, 30, 35, 40, or 45 amino acid residues and spans the plasma membrane. Transmembrane domains are rich in hydrophobic residues, and typically have an α-helical structure. In a preferred embodiment, at least 50%, 60%, 70%, 80%, 90%, 95% or more of the amino acids of a transmembrane domain are hydrophobic, *e.g.*, leucines, isoleucines, tyrosines, or tryptophans. Transmembrane domains are described in, for example, http://pfam.wustl.edu/cgi-bin/getdesc?name=7tm-1 and Zagotta et al, 1996, *Annual Rev. Neuronsci.* 19: 235-63, the contents of which are incorporated

herein by reference. Amino acid residues 375-398 of SEQ ID NO:5 (SEQ ID NO:104), 74-96 of SEQ ID NO:8 (SEQ ID NO:108), 768-791 of SEQ ID NO:14 (SEQ ID NO:115), 217-240 of SEQ ID NO:17 (SEQ ID NO:119), 501-524 of SEQ ID NO:20 (SEQ ID NO:123); 170-193 of SEQ ID NO:26 (SEQ ID NO:130), and 245-269, 287-306, 323-343, 358-376, 414-438, 457-477 and 485-504 of SEQ ID NO:29 (SEQ ID NOs:135-141) include transmembrane domains.

A MANGO 003, TANGO 272, TANGO 354 or TANGO 378 family member can include a C-terminal cytoplasmic domain. As used herein, a "C-terminal cytoplasmic domain" includes an amino acid sequence having a length of at least about 10, preferably about 10-25, more preferably about 25-50, more preferably about 50-75, even more preferably about 75-100, 100-133, 133-150, 150-200, 200-250, 250-300, 300-400, 400-500, or 500-600 amino acid residues and is located within a cell or within the cytoplasm of a cell. Accordingly, the N-terminal amino acid residue of a "C-terminal cytoplasmic domain" is adjacent to a C-terminal amino acid residue of a transmembrane domain in a naturally-occurring MANGO 003, TANGO 272, TANGO 354 or TANGO 378 protein. For example, a C-terminal cytoplasmic domain is found at about amino acid residues 399-504 of SEQ ID NO:5, 97-208 of SEQ ID NO:8, 792-1050 of SEQ ID NO:14, 241-497 of SEQ ID NO:17, 525-636 of SEQ ID NO:20; 194-305 of SEQ ID NO:26, and 505-528 of SEQ ID NO:29.

MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, or TANGO 378 family members can include a signal peptide. As used herein, a "signal peptide" includes a peptide of at least about 15 amino acid residues in length which occurs at the N-terminus of secretory and membrane-bound proteins and which contains at least about 70% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. The sequence can contain about 15 to 45 amino acid residues or about 17-22 amino acid residues, and has at least about 60-80%, 65-75%, or about 70% hydrophobic residues. A signal peptide serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a MANGO 003 protein contains a signal peptide of about amino acids 1-22, 1-23, 1-24, 1-25, or 1-26 of SEQ ID NO:5 (SEQ ID NO:101). In one embodiment, a MANGO 347 protein contains a signal peptide of about amino acids 1-33, 1-34, 1-35, 1-36, or 1-37 of SEQ ID NO:11 (SEQ ID NO:110). In one embodiment, a TANGO 272 protein contains a signal peptide of amino acids 1-18, 1-19, 1-20, 1-21, or 1-22 of SEQ ID NO:14 (SEQ ID NO:112). In yet another embodiment, a TANGO 295 protein contains a signal peptide of amino acids 1-26, 1-27, 1-28, 1-29, or 1-30 of SEQ ID NO:23 (SEQ ID NO:125). In another embodiment, a TANGO 354 protein contains a signal peptide of amino acids 1-17, 1-18, 1-19, 1-20, or 1-21 of SEQ ID NO:26 (SEQ ID NO:127). In another embodiment, a TANGO 378 protein contains a signal peptide of amino acids 1-19, 1-20, 1-21, 1-22, or 1-23 of SEQ ID NO:29 (SEQ ID

NO:132). The signal peptide is cleaved during processing of the mature protein. The amino acid sequence of the mature MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, or TANGO 378 protein starts at the next amino acid after the signal peptide is cleaved. For example, the amino acid sequence of MANGO 003 may start at amino acids 23, 24, 25, 26, or 27 depending on the exact location of the cleavage of the signal peptide.

The signal peptide is cleaved during processing of the mature protein. Sometimes the initial methionine residue is also cleaved from the protein during signal peptide processing. Thus, in one embodiment, a MANGO 003 protein does not contain a signal peptide or an initial methionine residue and begins from residue 2 of SEQ ID NO:102. In one embodiment, a MANGO 347 protein does not contain a signal peptide or an initial methionine residue and begins from residue 2 of SEQ ID NO:111. In one embodiment, a TANGO 272 protein does not contain a signal peptide or an initial methionine residue and begins from residue 2 of SEQ ID NO:113. Thus, in one embodiment, a TANGO 295 protein does not contain a signal peptide or an initial methionine residue an begins from residue 2 of SEQ ID NO:126. Thus, in one embodiment, a TANGO 354 protein does not contain a signal peptide or an initial methionine residue an begins from residue 2 of SEQ ID NO:128. Thus, in one embodiment, a TANGO 378 protein does not contain a signal peptide or an initial methionine residue an begins from residue 2 of SEQ ID NO:133.

In one embodiment, a MANGO 003 family member includes three immunoglobulin domains and a neurotransmitter-gated ion channel domain. In another embodiment, a MANGO 003 family member includes three immunoglobulin domains, a neurotransmitter-gated ion channel domain and a transmembrane domain. In yet another embodiment, a MANGO 003 family member includes three immunoglobulin domains, a neurotransmitter-gated ion channel domain, a transmembrane domain and an N-terminal extracellular domain. In another embodiment, a MANGO 003 family member includes three immunoglobulin domains, a neurotransmitter-gated ion channel domain, a transmembrane domain, an N-terminal extracellular domain and a C-terminal cytoplasmic domain. In yet another embodiment, a MANGO 003 family member includes three immunoglobulin domains, a neurotransmitter-gated ion channel domain, a transmembrane domain, an N-terminal extracellular domain, a C-terminal cytoplasmic domain, and a signal peptide.

In one embodiment, a MANGO 354 family member includes at least one immunoglobulin domain and a transmembrane domain. In another embodiment, a MANGO 354 family member includes at least one immunoglobulin domain, a transmembrane domain and a signal peptide.

In one embodiment, a TANGO 272 family member includes fourteen EGF-like domains and a delta serrate ligand domain. In another embodiment, a TANGO 272 family

member includes fourteen EGF-like domains, a delta serrate ligand domain and an RGD cell attachment site. In yet another embodiment, a TANGO 272 family member includes fourteen EGF-like domains, a delta serrate ligand domain, an RGD cell attachment site, and a transmembrane domain. In another embodiment, a TANGO 272 family member includes fourteen EGF-like domains, a delta serrate ligand domain, an RGD cell attachment site, a transmembrane domain, and an extracellular N-terminal domain. In another embodiment, a TANGO 272 family member includes fourteen EGF-like domains, a delta serrate ligand domain, an RGD cell attachment site, a transmembrane domain, an extracellular N-terminal domain and a C-terminal cytoplasmic domain. In another embodiment, a TANGO 272 family member includes fourteen EGF-like domains, a delta serrate ligand domain, an RGD cell attachment site, a transmembrane domain, an extracellular N-terminal domain, a C-terminal cytoplasmic domain, and a signal peptide.

In one embodiment, a TANGO 378 family member includes a 7 transmembrane receptor profile and three extracellular loops. In another embodiment, a TANGO 378 family member includes a 7 transmembrane receptor profile, three extracellular loops, and three cytoplasmic loops. In yet another embodiment, a TANGO 378 family member includes a 7 transmembrane receptor profile, three extracellular loops, three cytoplasmic loops, and an extracellular N-terminal domain. In another embodiment, a TANGO 378 family member includes a 7 transmembrane receptor profile, three extracellular loops, three cytoplasmic loops, an extracellular N-terminal domain, and a C-terminal cytoplasmic domain. In another embodiment, a TANGO 378 family member includes a 7 transmembrane receptor profile, three extracellular loops, three cytoplasmic loops, an extracellular N-terminal domain, a C-terminal cytoplasmic domain, and a signal peptide.

Various features of INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378 are summarized below.

INTERCEPT 340

A cDNA encoding INTERCEPT 340 was identified by analyzing the sequences of clones present in a human fetal spleen cDNA library.

This analysis led to the identification of a clone, jthsa102b12, encoding full-length human INTERCEPT 340. The cDNA of this clone is 3284 nucleotides long (Figures 1A-1B; SEQ ID NO:1). The 723 nucleotide open reading frame of this cDNA, nucleotides 1222-1944 of SEQ ID NO:1 (SEQ ID NO:3), encodes a 241 amino acid protein (Figures 1A-1B; SEQ ID NO:2).

Human INTERCEPT 340 that has not been post-translationally modified is predicted to have a molecular weight of 27.2 kDa.

Human INTERCEPT 340 includes three fibrillar collagen C-terminal (COLF) domains at amino acids 58-116 of SEQ ID NO:2 (SEQ ID NO:34); amino acids 126-151 of SEQ ID NO:2 (SEQ ID NO:35); and amino acids 186-217 of SEQ ID NO:2 (SEQ ID NO:36). Figure 3 depicts alignments of each of the COLF domains of human INTERCEPT 340 with consensus hidden Markov model COLF domains (SEQ ID NOs:31, 32, and 33). In one embodiment, INTERCEPT 340 is a secreted protein. In another embodiment, INTERCEPT 340 is a membrane-associated protein.

An N-glycosylation site is present at amino acids 105-108 of SEQ ID NO:2. A glycosaminoaglycan attachment site is present at amino acids 161-164 of SEQ ID NO:2. Protein kinase C phosphorylation sites are present at amino acids 57-59, 152-154, and 227-229 of SEQ ID NO:2. A tyrosine kinase phosphorylation site is present at amino acids 81-87 of SEQ ID NO:2. Casein kinase II phosphorylation sites are present at amino acids 36-39, 120-123 and 181-184. N-myristylation sites are present at amino acids 109-114 and 164-169 of SEQ ID NO:2.

Clone jthsa102b12, which encodes human INTERCEPT 340, was deposited as a composite deposit having a designation EpI340 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on June 18, 1999 and assigned Accession Number PTA-250. A description of the deposit conditions is set forth in the section entitled "Deposit of Clones" below. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 2 depicts a hydropathy plot of human INTERCEPT 340. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace.

Use of INTERCEPT 340 Nucleic Acids, Polypeptides, and Modulators Thereof

INTERCEPT 340 includes three fibrillar collagen C-terminal domains. Proteins having such domains play a role in modulating connective tissue formation and/or maintenance, and thus can influence a wide variety of biological processes, including assembly into fibrils; strengthening and organization of the extracellular matrix; shaping of tissues and cells; modulation of cell migration; and/or modulation of signal transduction pathways. Because INTERCEPT 340 includes fibrillar collagen C-terminal domains, INTERCEPT 340 polypeptides, nucleic acids, and modulators thereof can be used to treat connective tissue disorders, including a skin disorder and/or a skeletal disorder (*e.g.*, Marfan

syndrome and osteogenesis imperfecta); cardiovascular disorders including hyperproliferative vascular diseases (*e.g.*, hypertension, vascular restenosis and atherosclerosis), ischemia reperfusion injury, cardiac hypertrophy, coronary artery disease, myocardial infarction, arrhythmia, cardiomyopathies, and congestive heart failure); and/or hematopoietic disorders (*e.g.*, myeloid disorders, lymphoid malignancies, T cell disorders).

As INTERCEPT 340 was originally found in a fetal spleen library, INTERCEPT 340 nucleic acids, proteins, and modulators thereof can be used to modulate the function, survival, morphology, migration, proliferation and/or differentiation of cells that form the spleen, *e.g.*, cells of the splenic connective tissue, *e.g.*, splenic smooth muscle cells and/or endothelial cells of the splenic blood vessels. INTERCEPT 340 nucleic acids, proteins, and modulators thereof can also be used to modulate the proliferation, differentiation, and/or function of cells that are processed, *e.g.*, regenerated or phagocytized within the spleen, *e.g.*, erythrocytes and/or B and T lymphocytes and macrophages. Thus INTERCEPT 340 nucleic acids, proteins, and modulators thereof can be used to treat spleen, *e.g.*, the fetal spleen, associated diseases and disorders. Examples of splenic diseases and disorders include *e.g.*, splenic lymphoma and/or splenomegaly, and/or phagocytotic disorders, *e.g.*, those inhibiting macrophage engulfment of bacteria and viruses in the bloodstream.

Further, in light of INTERCEPT 340's presence in a human fetal spleen cDNA library, INTERCEPT 340 expression can be utilized as a marker for specific tissues (*e.g.*, lymphoid tissues such as the spleen) and/or cells (*e.g.*, splenic) in which INTERCEPT 340 is expressed. INTERCEPT 340 nucleic acids can also be utilized for chromosomal mapping.

<u>MANGO 003</u>

A cDNA encoding human MANGO 003 was identified by analyzing the sequences of clones present in a human thyroid cDNA library.

This analysis led to the identification of a clone, jthYa030d03, encoding full-length human MANGO 003. The cDNA of this clone is 3169 nucleotides long (Figures 4A-4B; SEQ ID NO:4). The 1512 nucleotide open reading frame of this cDNA, nucleotide 57 to nucleotide 1568 of SEQ ID NO:4 (SEQ ID NO:6), encodes a 504 amino acid protein (Figures 4A-4B; SEQ ID NO:5).

Human MANGO 003 that has not been post-translationally modified is predicted to have a molecular weight of 54.5 kDa prior to cleavage of its signal peptide (52.1 kDa after cleavage of its signal peptide).

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human MANGO 003 includes a 24 amino acid signal peptide at amino acid 1 to about amino acid 24 of SEQ ID NO:5 (SEQ ID NO:101) preceding the mature human MANGO 003 protein which corresponds to about amino acid 25 to amino acid 504 of SEQ ID NO:5 (SEQ ID NO:102).

Human MANGO 003 is a transmembrane protein having an extracellular domain which extends from about amino acid 25 to about amino acid 374 of SEQ ID NO:5 (SEQ ID NO:103), a transmembrane domain which extends from about amino acid 375 to about amino acid 398 of SEQ ID NO:5 (SEQ ID NO:104), and a cytoplasmic domain which extends from about amino acid 399 to amino acid 504 of SEQ ID NO:5 (SEQ ID NO:105).

Alternatively, in another embodiment, a human MANGO 003 protein contains an extracellular domain which extends from about amino acid 399 to amino acid 504 of SEQ ID NO:5 (SEQ ID NO:105), a transmembrane domain which extends from about amino acid 375 to about amino acid 398 of SEQ ID NO:5 (SEQ ID NO:104), and a cytoplasmic domain which extends from about amino acid 25 to about amino acid 374 of SEQ ID NO:5 (SEQ ID NO:103).

Human MANGO 003 includes three immunoglobulin domains at amino acids 44-101 of SEQ ID NO:5 (SEQ ID NO:38); amino acids 165-223 of SEQ ID NO:5 (SEQ ID NO:39); and amino acids 261-340 of SEQ ID NO:5 (SEQ ID NO:40). Figure 6 depicts alignments of each of the immunoglobulin domains of MANGO 003 with a consensus hidden Markov model immunoglobulin domain (SEQ ID NO:37).

Human MANGO 003 includes a neurotransmitter gated ion channel domain at amino acids 388-397 of SEQ ID NO:5 (SEQ ID NO:43). Figure 7 depicts an alignment of the neurotransmitter gated ion channel domain of human MANGO 003 with a neurotransmitter gated ion channel domain derived from a hidden Markov model (SEQ ID NO:42).

N-glycosylation sites are present at amino acids 111-114, 231-234, 255-258, and 293-296 of SEQ ID NO:5. A cAMP and cGMP-dependent protein kinase phosphorylation site is present at amino acids 202-205 of SEQ ID NO:5. Protein kinase C phosphorylation sites are present at amino acids 44-48, 167-169, 207-209, 216-218, 220-222, 224-226, 233-235, 347-349, and 422-424 of SEQ ID NO:5. Casein kinase II phosphorylation sites are present at amino acids 192-195, 256-259, 294-297, 313-316, 422-425, and 490-493 of SEQ ID NO:5. Tyrosine kinase phosphorylation sites are present at amino acids 212-219 and 329-336 of SEQ ID NO:5. N-myristylation sites are present at amino acids 95-100, 228-233, 261-266, 317-322, 334-339, 382-387, and 443-448 of SEQ ID NO:5.

Clone jthYa030d03, which encodes human MANGO 003, was deposited as a composite deposit having a designation EpthLa6a1 with the American Type Culture

Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on March 27, 1999 and assigned Accession Number 207178. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 5 depicts a hydropathy plot of human MANGO 003. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 5 indicates the presence of a hydrophobic domain within human MANGO 003, suggesting that human MANGO 003 is a transmembrane protein.

A cDNA encoding mouse MANGO 003 was identified by analyzing the sequences of clones present in a mouse choroid plexus cDNA library.

This analysis led to the identification of a clone, jfmjf004c11, encoding partial mouse MANGO 003. The cDNA of this clone is 504 nucleotides long (Figures 8A-8B; SEQ ID NO:7). The 626 nucleotide open reading frame of this cDNA, nucleotides 1-626 of SEQ ID NO:7 (SEQ ID NO:9), encodes a 208 amino acid protein (Figures 8A-8B; SEQ ID NO:8).

Northern blot analysis using the mouse clone jfmjf004c11 revealed strong expression of the mouse MANGO 003 gene in the mouse liver, skeletal muscle and kidney. Moderate expression was detected in the heart, lung and testis, and lower levels of expression were detected in the mouse brain. No expression was detected in the spleen.

Mouse MANGO 003 that has not been post-translationally modified is predicted to have a molecular weight of 22.3 kDa.

Mouse MANGO 003 is a transmembrane protein having an extracellular domain which extends from about amino acid 1 to about amino acid 73 of SEQ ID NO:8 (SEQ ID NO:107), a transmembrane domain which extends from about amino acid 74 to about amino acid 96 of SEQ ID NO:8 (SEQ ID NO:108), and a cytoplasmic domain which extends from about amino acid 97 to amino acid 208 of SEQ ID NO:8 (SEQ ID NO:109).

An N-glycosylation site is present at amino acids 190-193 of SEQ ID NO:8. Protein kinase C phosphorylation sites are present at amino acids 44-46, 98-100, 119-121, and 197-199 of SEQ ID NO:8. Casein kinase II phosphorylation sites are present at amino acids 10-13, and 119-122 of SEQ ID NO:8. A tyrosine kinase phosphorylation site is present at amino acids 26-33 of SEQ ID NO:8. N-myristylation sites are present at amino acids 14-19, 31-36, and 79-84 of SEQ ID NO:8.

Figure 9 depicts a hydropathy plot of mouse MANGO 003. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 9 indicates the presence of a hydrophobic domain within human MANGO 003, suggesting that human MANGO 003 is a transmembrane protein.

A global alignment between the nucleotide sequence of the open reading frame (ORF) of human MANGO 003 (SEQ ID NO:6) and the nucleotide sequence of the open reading frame of mouse MANGO 003 (SEQ ID NO:9) revealed a 31.1% identity (Figures 27A-27C). The global alignment was performed using the ALIGN program version 2.0u (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of –1212; Myers and Miller, 1989 *CABIOS* 4:11-7).

A local alignment between the nucleotide sequence of human MANGO 003 (SEQ ID NO:4) and the nucleotide sequence of mouse MANGO 003 (SEQ ID NO:7) revealed a 62.8 % identity over nucleotides 970-2080 of the human MANGO 003 sequence (nucleotides 10-1070 of mouse MANGO 003) (Figures 28A-28B). The local alignment was performed using the L-ALIGN program version 2.0u54 July 1996 (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a score of 3241; Huang and Miller, 1991, *Adv. Appl. Math.* 12:373-81).

A global alignment between the amino acid sequence of human MANGO 003 (SEQ ID NO:5) and the amino acid sequence of mouse MANGO 003 (SEQ ID NO:8) revealed a 30.1% identity (Figure 29). The global alignment was performed using the ALIGN program version 2.0u (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of –488; Myers and Miller, 1989, *CABIOS* 4:11-7).

Use of MANGO 003 Nucleic Acids, Polypeptides, and Modulators Thereof

MANGO 003 includes three immunoglobulin-like domains. Proteins having such domains play a role in mediating protein-protein and protein-ligand interactions, and thus can influence a wide variety of biological processes, including cell surface recognition; transduction of an extracellular signal (*e.g.*, by interacting with a ligand and/or a cell-surface receptor); and/or modulation of signal transduction pathways.

MANGO 003 further includes a neurotransmitter-gated ion channel domain. Proteins having such domains play a role in modulating signal transmission at chemical synapses by, for example, influencing processes, such as the release of neurotransmitters from a cell (*e.g.*, a neuronal cell); modulating membrane excitability and/or resting potential; and/or modulating ion flux across a membrane of a cell (*e.g.*, a neuronal or a muscle cell). Because MANGO 003 includes a neurotransmitter-gated ion channel domain,

MANGO 003 polypeptides, nucleic acids, and modulators thereof can be used to treat neural disorders (*e.g.*, a CNS disorder, including Alzheimer's disease, Pick's disease, Parkinson's and other Lewy diffuse body diseases, multiple sclerosis, amyotrophic lateral sclerosis, progressive supranuclear palsy, epilepsy, and Jakob-Creutzfieldt disease; psychiatric disorders, *e.g.*, depression, schizophrenic disorders, Korsakoff's psychosis, mania, anxiety disorders, or phobic disorders; learning or memory disorders, *e.g.*, amnesia or age-related memory loss; and neurological disorders, *e.g.*, migraine).

MANGO 003 polypeptides, nucleic acids, and modulators thereof can be used to modulate function, survival, morphology, migration, proliferation and/or differentiation of cells in the tissues in which it is expressed (*e.g.* thyroid, liver, skeletal muscle, kidney, heart, lung, testis and brain). For example, MANGO 003 polypeptides, nucleic acids, and modulators thereof can be used to modulate endocrine, hepatic, skeletal muscular, renal, cardiac, reproductive and/or brain function. Accordingly, these molecules can be used to treat a variety of disease including, but not limited to, endocrine disorders (*e.g.*, hypothyroidism, hyperthyroidism, dwarfism, giantism, acromegaly); hepatic disorders (*e.g.*, hepatitis, liver cirrhosis, hepatoma, liver cysts, and hepatic vein thrombosis); skeletal muscular disorders; renal disorders (*e.g.*, renal cell carcinoma, nephritis, polycystic kidney disease); cardiovascular disorders (*e.g.*, atherosclerosis, ischemia reperfusion injury, cardiac hypertrophy, hypertension, coronary artery disease, myocardial infarction, arrhythmia, cardiomyopathies, and congestive heart failure); and/or reproductive disorders (*e.g.*, sterility).

MANGO 003 polypeptides, nucleic acids, or modulators thereof, can be used to treat hepatic (liver) disorders, such as jaundice, hepatic failure, hereditary hyperbiliruinemias (*e.g.*, Gilbert's syndrome, Crigler-Naijar syndromes and Dubin-Johnson and Rotor's syndromes), hepatic circulatory disorders (*e.g.*, hepatic vein thrombosis and portal vein obstruction and thrombosis) hepatitis (*e.g.*, chronic active hepatitis, acute viral hepatitis, and toxic and drug-induced hepatitis) cirrhosis (*e.g.*, alcoholic cirrhosis, biliary cirrhosis, and hemochromatosis), or malignant tumors (*e.g.*, primary carcinoma, hepatoblastoma, and angiosarcoma).

In another example, MANGO 003 polypeptides, nucleic acids, or modulators thereof, can be used to treat disorders of skeletal muscle, such as muscular dystrophy (*e.g.*, Duchenne Muscular Dystrophy, Becker Muscular Dystrophy, Emery-Dreifuss Muscular Dystrophy, Limb-Girdle Muscular Dystrophy, Facioscapulohumeral Muscular Dystrophy, Myotonic Dystrophy, Oculopharyngeal Muscular Dystrophy, Distal Muscular Dystrophy, and Congenital Muscular Dystrophy), motor neuron diseases (*e.g.*, Amyotrophic Lateral Sclerosis, Infantile Progressive Spinal Muscular Atrophy, Intermediate Spinal Muscular Atrophy, Spinal Bulbar Muscular Atrophy, and Adult Spinal Muscular Atrophy),

myopathies (*e.g.*, inflammatory myopathies (*e.g.*, Dermatomyositis and Polymyositis), Myotonia Congenita, Paramyotonia Congenita, Central Core Disease, Nemaline Myopathy, Myotubular Myopathy, and Periodic Paralysis), and metabolic diseases of muscle (*e.g.*, Phosphorylase Deficiency, Acid Maltase Deficiency, Phosphofructokinase Deficiency, Debrancher Enzyme Deficiency, Mitochondrial Myopathy, Carnitine Deficiency, Carnitine Palmityl Transferase Deficiency, Phosphoglycerate Kinase Deficiency, Phosphoglycerate Mutase Deficiency, Lactate Dehydrogenase Deficiency, and Myoadenylate Deaminase Deficiency).

In another example, MANGO 003 polypeptides, nucleic acids, or modulators thereof, can be used to treat renal disorders, such as glomerular diseases (*e.g.*, acute and chronic glomerulonephritis, rapidly progressive glomerulonephritis, nephrotic syndrome, focal proliferative glomerulonephritis, glomerular lesions associated with systemic disease, such as systemic lupus erythematosus, Goodpasture's syndrome, multiple myeloma, diabetes, neoplasia, sickle cell disease, and chronic inflammatory diseases), tubular diseases (*e.g.*, acute tubular necrosis and acute renal failure, polycystic renal diseasemedullary sponge kidney, medullary cystic disease, nephrogenic diabetes, and renal tubular acidosis), tubulointerstitial diseases (*e.g.*, pyelonephritis, drug and toxin induced tubulointerstitial nephritis, hypercalcemic nephropathy, and hypokalemic nephropathy) acute and rapidly progressive renal failure, chronic renal failure, nephrolithiasis, vascular diseases (*e.g.*, hypertension and nephrosclerosis, microangiopathic hemolytic anemia, atheroembolic renal disease, diffuse cortical necrosis, and renal infarcts), or tumors (*e.g.*, renal cell carcinoma and nephroblastoma).

Further, in light of MANGO 003's pattern of expression in mice, MANGO 003 expression can be utilized as a marker for specific tissues (*e.g.*, liver, skeletal muscle, kidney) and/or cells (*e.g.*, hepatic, skeletal muscle, renal) in which MANGO 003 is expressed. MANGO 003 nucleic acids can also be utilized for chromosomal mapping.

<u>MANGO 347</u>

A cDNA encoding human MANGO 347 was identified by analyzing the sequences of clones present in a human brain cDNA library.

This analysis led to the identification of a clone, jlhbad295g12, encoding full-length human MANGO 347. The cDNA of this clone is 1423 nucleotides long (Figure 10; SEQ ID NO:10). The 414 nucleotide open reading frame of this cDNA, nucleotides 31 to 444 of

SEQ ID NO:10 (SEQ ID NO:12), encodes a 138 amino acid protein (Figure 10; SEQ ID NO:11).

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human MANGO 347 includes a 35 amino acid signal peptide at amino acid 1 to about amino acid 35 of SEQ ID NO:11 (SEQ ID NO:110) preceding the mature human MANGO 347 protein which corresponds to about amino acid 36 to amino acid 138 of SEQ ID NO:11 (SEQ ID NO:111).

Human MANGO 347 that has not been post-translationally modified is predicted to have a molecular weight of 15.4 kDa prior to cleavage of its signal peptide and a molecular weight of 11.3 kDa subsequent to cleavage of its signal peptide.

Human MANGO 347 includes a CUB domain at amino acids 40-136 of SEQ ID NO:11 (SEQ ID NO:45). An alignment of the CUB domain of human MANGO 347 with a consensus hidden Markov model CUB domain amino acid sequence derived from a hidden Markov model (SEQ ID NO:44) is shown in Figure 12.

Casein kinase II phosphorylation sites are present at amino acids 67-70, and 108-111 of SEQ ID NO:11. N-myristylation sites are present at amino acids 19-24, 31-36, 64-69, and 113-118 of SEQ ID NO:11.

Clone jlhbad295g12, which encodes human MANGO 347, was deposited as a composite deposit having a designation EpM347 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on June 18, 1999 and assigned Accession Number PTA-250. A description of the deposit conditions used in set forth below. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 11 depicts a hydropathy plot of human MANGO 347. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 11 indicates that human MANGO 347 has a signal peptide at its amino terminus, suggesting that human MANGO 347 is a secreted protein.

Use of MANGO 347 Nucleic Acids, Polypeptides, and Modulators Thereof

MANGO 347 includes a CUB domain. Proteins having such a domain play a role in mediating cell interactions during development, and thus can influence a wide variety of developmental processes, including morphogenesis, cellular migration, adhesion, proliferation, differentiation, and/or survival. MANGO 347 polypeptides are expressed in

neural (*e.g.*, brain cells). Because MANGO 347 includes a CUB domain and is expressed in neural cells, MANGO 347 polypeptides, nucleic acids, and modulators thereof can be used to treat disorders involving, *e.g.*, cellular migration, proliferation, and differentiation of a cell, *e.g.*, a neural cell (*e.g.*, a CNS disorder, including Alzheimer's disease, Pick's disease, Parkinson's and other Lewy diffuse body diseases, multiple sclerosis, amyotrophic lateral sclerosis, progressive supranuclear palsy, epilepsy, and Jakob-Creutzfieldt disease; psychiatric disorders, *e.g.*, depression, schizophrenic disorders, Korsakoff's psychosis, mania, anxiety disorders, or phobic disorders; learning or memory disorders, *e.g.*, amnesia or age-related memory loss; and neurological disorders, *e.g.*, migraine).

Further, in light of MANGO 347's presence in a human brain cDNA library, MANGO 347 expression can be utilized as a marker for specific tissues (*e.g.*, brain) and/or cells (*e.g.*, brain) in which MANGO 347 is expressed. MANGO 347 nucleic acids can also be utilized for chromosomal mapping.

<u>TANGO 272</u>

A cDNA encoding human TANGO 272 was identified by analyzing the sequences of clones present in a human microvascular endothelial cell library (HMVEC) cDNA library.

This analysis led to the identification of a clone, jthda089h03, encoding full-length human TANGO 272. The cDNA of this clone is 5036 nucleotides long (Figures 13A-13D; SEQ ID NO:13). The 3149 nucleotide open reading frame of this cDNA, nucleotides 230-3379 of SEQ ID NO:13 (SEQ ID NO:15), encodes a 1050 amino acid protein (Figures 13A-13D; SEQ ID NO:14).

Northern blot analysis using the human clone jthda089h03 revealed strong expression of the human TANGO 272 gene in the heart. Moderate expression was detected in the placenta, lung, and liver, and lower levels of expression were detected in the brain, skeletal muscle, kidney, and pancreas.

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human TANGO 272 includes an 20 amino acid signal peptide at amino acid 1 to about amino acid 20 of SEQ ID NO:14 (SEQ ID NO:112) preceding the mature human TANGO 272 protein which corresponds to about amino acid 21 to amino acid 1050 of SEQ ID NO:14 (SEQ ID NO:113).

Human TANGO 272 that has not been post-translationally modified is predicted to have a molecular weight of 112 kDa prior to cleavage of its signal peptide and a molecular weight of 110 kDa subsequent to cleavage of its signal peptide.

Human TANGO 272 is a transmembrane protein having an extracellular domain which extends from about amino acid 21 to about amino acid 767 of SEQ ID NO:14 (SEQ

ID NO:114), a transmembrane domain which extends from about amino acid 768 to about amino acid 791 of SEQ ID NO:14 (SEQ ID NO:115), and a cytoplasmic domain which extends from about amino acid 792 to amino acid 1050 of SEQ ID NO:14 (SEQ ID NO:116).

Alternatively, in another embodiment, a human TANGO 272 protein contains an extracellular domain which extends from about amino acid 792 to amino acid 1050 of SEQ ID NO:14 (SEQ ID NO:116), a transmembrane domain which extends from about amino acid 768 to about amino acid 791 of SEQ ID NO:14 (SEQ ID NO:115), and a cytoplasmic domain which extends from about amino acid 21 to about amino acid 767 of SEQ ID NO:14 (SEQ ID NO:114).

Human TANGO 272 includes fourteen EGF-like domains at amino acids 151-181 of SEQ ID NO:14 (SEQ ID NO:49); amino acids 200-229 of SEQ ID NO:14 (SEQ ID NO:50); amino acids 242-272 of SEQ ID NO:14 (SEQ ID NO:51); amino acids 285-315 of SEQ ID NO:14 (SEQ ID NO:52); amino acids 328-358 of SEQ ID NO:14 (SEQ ID NO:53); amino acids 378-404 of SEQ ID NO:14 (SEQ ID NO:54); amino acids 417-447 of SEQ ID NO:14 (SEQ ID NO:55); amino acids 460-490 of SEQ ID NO:14 (SEQ ID NO:56); amino acids 503-533 of SEQ ID NO:14 (SEQ ID NO:57); amino acids 546-576 of SEQ ID NO:14 (SEQ ID NO:58); amino acids 589-619 of SEQ ID NO:14 (SEQ ID NO:59); amino acids 632-661 of SEQ ID NO:14 (SEQ ID NO:60); amino acids 674-704 of SEQ ID NO:14 (SEQ ID NO:61); and amino acids 717-747 of SEQ ID NO:14 (SEQ ID NO:62). Figures 15A-15C depict alignments of each of the EGF-like domains of TANGO 272 with consensus hidden Markov model EGF-like domains (SEQ ID NO:46). Human TANGO 272 further includes a delta serrate ligand domain from amino acids 518 to 576 of SEQ ID NO:14 (SEQ ID NO:63). An alignment of the delta serrate ligand domain of human TANGO 272 with a consensus hidden Markov model of this domain (SEQ ID NO:47) is also depicted (Figure 15B).

An RGD cell attachment site is present at amino acids 177-179 of SEQ ID NO:14. N-glycosylation sites are present at amino acids 284-287, 405-408, 459-462, 489-492, 504-507, 588-591, 639-642, 647-650, 716-719, and 873-876 of SEQ ID NO:14. An amidation site is present at amino acids 628-631 of SEQ ID NO:14. Protein kinase C phosphorylation sites are present at amino acids 38-40, 70-72, 107-109, 359-361, 461-463, 594-596, 809-811, 896-898, 940-942, 977-979, and 1022-1024 of SEQ ID NO:14. Casein kinase II phosphorylation sites are present at amino acids 30-33, 38-41, 473-476, 548-551, 579-582, 657-660, 897-900, 921-924, 940-943, and 955-958 of SEQ ID NO:14. A tyrosine kinase phosphorylation site is present at amino acids 361-368 of SEQ ID NO:14. N-myristylation sites are present at amino acids 14-19, 103-108, 269-274, 302-307, 325-330, 345-350, 401-

406, 427-432, 434-439, 457-462, 520-525, 586-591, 606-611, 648-653, 707-712, 714-719, 769-774, 866-871, 926-931, and 1014-1019 of SEQ ID NO:14.

Clone jthda089h03, which encodes human TANGO 272, was deposited as a composite deposit having a designation EpT272 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2236) June 18, 1999 and assigned Accession Number PTA-250. A description of the deposit conditions used is set forth in the section entitled "Deposit of Clones" below. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 14 depicts a hydropathy plot of human TANGO 272. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 16 indicates the presence of a hydrophobic domain within human TANGO 272, suggesting that human TANGO 272 is a transmembrane protein.

A cDNA encoding mouse TANGO 272 was identified by analyzing the sequences of clones present in a mouse testis cDNA library.

This analysis led to the identification of a clone, jtmzb062c04, encoding partial mouse TANGO 272. The cDNA of this clone is 2569 nucleotides long (Figures 16A-16B; SEQ ID NO:16). The 1492 nucleotide open reading frame of this cDNA, nucleotides 1-1492 of SEQ ID NO:16 (SEQ ID NO:18), encodes a 497 amino acid protein (Figures 16A-16B; SEQ ID NO:17).

Mouse TANGO 272 that has not been post-translationally modified is predicted to have a molecular weight of 53.5 kDa.

Mouse TANGO 272 is a transmembrane protein having an extracellular domain which extends from about amino acid 1 to about amino acid 216 of SEQ ID NO:17 (SEQ ID NO:118), a transmembrane domain which extends from about amino acid 217 to about amino acid 240 of SEQ ID NO:17 (SEQ ID NO:119), and a cytoplasmic domain which extends from about amino acid 241 to amino acid 497 of SEQ ID NO:17 (SEQ ID NO:120).

Alternatively, in another embodiment, a mouse TANGO 272 protein contains an extracellular domain which extends from about amino acid 241 to amino acid 497 of SEQ ID NO:17 (SEQ ID NO:120), a transmembrane domain which extends from about amino acid 217 to about amino acid 240 of SEQ ID NO:17 (SEQ ID NO:119), and a cytoplasmic domain which extends from about amino acid 1 to about amino acid 216 of SEQ ID NO:17 (SEQ ID NO:118).

Mouse TANGO 272 includes four EGF-like domains at about amino acids 37-67 of SEQ ID NO:17 (SEQ ID NO:64); amino acids 80-110 of SEQ ID NO:17 (SEQ ID NO:65); amino acids 123-153 of SEQ ID NO:17 (SEQ ID NO:66); and amino acids 166-196 of SEQ ID NO:17 (SEQ ID NO:67). Mouse TANGO 272 further includes four laminin-EGF-like domains at about amino acids 3-37 of SEQ ID NO:17 (SEQ ID NO:68); amino acids 41-80 of SEQ ID NO:17 (SEQ ID NO:69); amino acids 83-123 of SEQ ID NO:17 (SEQ ID NO:70); and amino acids 127-172 of SEQ ID NO:17 (SEQ ID NO:71). Figures 39A-39B depict alignments of each of the EGF-like- and laminin-EGF-like domains of TANGO 272 with consensus hidden Markov model EGF-like domains (SEQ ID NOs:46 and 48, respectively).

Mouse TANGO 272 further includes a delta serrate ligand domain from amino acids 10 to 67 of SEQ ID NO:17 (SEQ ID NO:72). An alignment of the delta serrate ligand domain of mouse TANGO 272 with a consensus hidden Markov model of this domain (SEQ ID NO:47) is also depicted in Figures 39A-39B.

Based on the Prosite analysis, EGF-like domain cysteine pattern signature are present at amino acids 13-24, 56-67, 99-110, 142-153, and 185-196 of SEQ ID NO:17.

N-glycosylation sites are present at amino acids 36-39, 88-91, 165-168, and 323-326 of SEQ ID NO:17. An amidation site is present at amino acids 76-79 of SEQ ID NO:17. Protein kinase C phosphorylation sites are present at amino acids 42-44, 258-260, 354-356, 388-390, 469-471, and 492-494 of SEQ ID NO:17. Casein kinase II phosphorylation sites are present at amino acids 106-109, 192-195, 343-346, 388-391, and 446-449 of SEQ ID NO:17. N-myristylation sites are present at amino acids 11-16, 34-39, 47-52, 54-59, 97-102, 120-125, 140-145, 163-168, 199-204, 218-223, 372-377, and 461-466 of SEQ ID NO:17.

Figure 17 depicts a hydropathy plot of mouse TANGO 272. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 17 indicates the presence of a hydrophobic domain within mouse TANGO 272, suggesting that mouse TANGO 272 is a transmembrane protein.

A cDNA encoding rat TANGO 272 was identified by analyzing the sequences of clones present in a rat neonatal sciatic nerve cDNA library.

This analysis led to the identification of a clone, atrxa6b6, encoding partial rat TANGO 272. The cDNA of this clone is 3567 nucleotides long (Figures 33A-33C; SEQ ID NO:19). The 1908 nucleotide open reading frame of this cDNA, nucleotides 925-2832 of SEQ ID NO:19 (SEQ ID NO:21), encodes a 636 amino acid protein (Figures 33A-33C; SEQ ID NO:20).

Rat TANGO 272 that has not been post-translationally modified is predicted to have a molecular weight of 67.4 kDa.

Rat TANGO 272 is a transmembrane protein having an extracellular domain which extends from about amino acid 1 to about amino acid 500 of SEQ ID NO:20 (SEQ ID NO:122), a transmembrane domain which extends from about amino acid 501 to about amino acid 524 of SEQ ID NO:20 (SEQ ID NO:123), and a cytoplasmic domain which extends from about amino acid 525 to amino acid 636 of SEQ ID NO:20 (SEQ ID NO:124).

Alternatively, in another embodiment, a rat TANGO 272 protein contains an extracellular domain which extends from about amino acid 525 to amino acid 636 of SEQ ID NO:20 (SEQ ID NO:124), a transmembrane domain which extends from about amino acid 501 to about amino acid 524 of SEQ ID NO:20 (SEQ ID NO:123), and a cytoplasmic domain which extends from about amino acid 1 to about amino acid 500 of SEQ ID NO:20 (SEQ ID NO:122).

Rat TANGO 272 includes eleven EGF-like domains at about amino acids 18-48 of SEQ ID NO:20 (SEQ ID NO:73); amino acids 61-91 of SEQ ID NO:20 (SEQ ID NO:74); amino acids 105-137 of SEQ ID NO:20 (SEQ ID NO:75); amino acids 150-180 of SEQ ID NO:20 (SEQ ID NO:76); amino acids 193-223 of SEQ ID NO:20 (SEQ ID NO:77); amino acids 236-266 of SEQ ID NO:20 (SEQ ID NO:78); amino acids 279-309 of SEQ ID NO:20 (SEQ ID NO:79); amino acids 322-352 of SEQ ID NO:20 (SEQ ID NO:80); amino acids 365-394 of SEQ ID NO:20 (SEQ ID NO:81); amino acids 407-437 of SEQ ID NO:20 (SEQ ID NO:82); and amino acids 450-480 of SEQ ID NO:20 (SEQ ID NO:83). Figures 41A-41D depict alignments of each of the EGF-like-domains of rat TANGO 272 with consensus hidden Markov model EGF-like domains (SEQ ID NO:46).

Rat TANGO 272 further includes eleven laminin/EGF-like domains at about amino acids 22-61 of SEQ ID NO:20 (SEQ ID NO:84); amino acids 65-105 of SEQ ID NO:20 (SEQ ID NO:85); amino acids 109-150 of SEQ ID NO:20 (SEQ ID NO:86); amino acids 154-193 of SEQ ID NO:20 (SEQ ID NO:87); amino acids 197-236 of SEQ ID NO:20 (SEQ ID NO:88); amino acids 240-279 of SEQ ID NO:20 (SEQ ID NO:89); amino acids 283-322 of SEQ ID NO:20 (SEQ ID NO:90); amino acids 326-365 of SEQ ID NO:20 (SEQ ID NO:91); amino acids 368-407 of SEQ ID NO:20 (SEQ ID NO:92); amino acids 411-450; and amino acids 454-489 of SEQ ID NO:20 (SEQ ID NO:93). Figures 41A-41D depict alignments of each of the laminin/EGF-like-domains of rat TANGO 272 with consensus hidden Markov model EGF-like domains (SEQ ID NO:48).

Rat TANGO 272 further includes a delta serrate ligand domain from amino acids 246 to 309 of SEQ ID NO:20 (SEQ ID NO:95). An alignment of the delta serrate ligand domain of rat TANGO 272 with a consensus hidden Markov model of this domain (SEQ ID NO:47) is also depicted in Figures 41A-41D.

Based on the Prosite analysis, EGF-like domain cysteine pattern signature are present at amino acids 37-48, 80-91, 126-137, 169-180, 255-266, 298-309, 341-352, 383-394, 426-437, and 469-480 of SEQ ID NO:20.

N-glycosylation sites are present at amino acids 17-20, 138-141, 192-195, 222-225, 237-240, 321-324, 372-375, 436-439, and 449-452 of SEQ ID NO:20. A cAMP/cGMP-dependent protein kinase phosphorylation site is present at amino acids 618-621 of SEQ ID NO:20. An amidation site is present at amino acids 361-364 of SEQ ID NO:20. Protein kinase C phosphorylation sites are present at amino acids 92-94, 327-329, 542-544, and 596-598 of SEQ ID NO:20. Casein kinase II phosphorylation sites are present at amino acids 104-107, 206-209, 281-284, and 390-393 of SEQ ID NO:20. A tyrosine kinase phosphorylation site is present at amino acids 94-101 of SEQ ID NO:20. N-myristylation sites are present at amino acids 2-7, 35-40, 58-63, 78-83, 134-139, 160-165, 167-172, 190-195, 210-215, 253-258, 319-324, 339-344, 381-386, 404-409, 424-429, 447-452, 483-488, and 502-507 of SEQ ID NO:20.

Figure 40 depicts a hydropathy plot of rat TANGO 272. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 40 indicates the presence of a hydrophobic domain within rat TANGO 272, suggesting that rat TANGO 272 is a transmembrane protein.

A global alignment between the nucleotide sequence of the open reading frame (ORF) of human TANGO 272 (SEQ ID NO:15) and the nucleotide sequence of the open reading frame of mouse TANGO 272 (SEQ ID NO:18) revealed a 39.1% identity (Figures 30A-30E). The global alignment was performed using the ALIGN program version 2.0u (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of –79; Myers and Miller, 1989, *CABIOS* 4:11-7).

A local alignment between the nucleotide sequence of human TANGO 272 (SEQ ID NO:13) and the nucleotide sequence of mouse TANGO 272 (SEQ ID NO:16) revealed 67.6 % identity over nucleotides 1890-4610 of the human TANGO 272 sequence (nucleotides 10-2560 of mouse TANGO 272) (Figures 31A-31D). The local alignment was performed using the L-ALIGN program version 2.0u54 July 1996 (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a score of 8462; Huang and Miller, 1991, *Adv. Appl. Math.* 12:373-81).

A global alignment between the amino acid sequence of human TANGO 272 (SEQ ID NO:14) and the amino acid sequence of mouse TANGO 272 (SEQ ID NO:17) revealed a 38.2% identity (Figures 32A-32B). The global alignment was performed using the ALIGN

program version 2.0u (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of –19; Myers and Miller, 1989, *CABIOS* 4:11-7).

A global alignment between the nucleotide sequence of human TANGO 272 (SEQ ID NO:13) and the nucleotide sequence of rat TANGO 272 (SEQ ID NO:19) revealed a 55.7% identity (Figures 34A-34H). The global alignment was performed using the ALIGN program version 2.0u (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of 8635; Myers and Miller, 1989, *CABIOS* 4:11-7).

A global alignment between the nucleotide sequence of mouse TANGO 272 (SEQ ID NO:16) and the nucleotide sequence of rat TANGO 272 (SEQ ID NO:19) revealed a 43.7% identity (Figures 35A-35F). The global alignment was performed using the ALIGN program version 2.0u (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of 2827; Myers and Miller, 1989, *CABIOS* 4:11-7).

Use of TANGO 272 Nucleic Acids, Polypeptides, and Modulators Thereof

TANGO 272 includes fourteen EGF-like domains. Proteins having such domains play a role in mediating protein-protein interactions, and thus can influence a wide variety of biological processes, including cell surface recognition; modulation of cell-cell contact; modulation of cell fate determination; and modulation of wound healing and tissue repair.

TANGO 272 further includes an RGD cell attachment site. Proteins having such domains are typically extracellular matrix proteins such as collagens, laminin and fibronectin, among others (reviewed in Ruoslahti, 1996, *Annu. Rev. Cell Dev. Biol.* 12:697-715). An RGD cell attachment site typically interacts (*e.g.*, binds to) a cell surface receptor, such as an integrin receptor, and thus mediates a variety of biological processes, including cellular adhesion, migration, among others.

Because TANGO 272 includes EGF-like domains and an RGD cell attachment site, TANGO 272 polypeptides, nucleic acids, and modulators thereof can be used to treat disorders involving, *e.g.*, cellular migration, proliferation, and differentiation of a cell. For example, TANGO 272 polypeptides, nucleic acids, and modulators thereof can be used to treat neoplastic disorders, *e.g.*, cancer, tumor metastasis.

TANGO 272 polypeptides, nucleic acids, and modulators thereof can be used to modulate function, survival, morphology, migration, proliferation, tissue repair and/or differentiation of cells in the tissues in which it is expressed (*e.g.*, microvascular endothelial cells). For example, TANGO 272 polypeptides, nucleic acids, and modulators thereof can be used to modulate cardiovascular function, and/or to promote wound healing and tissue repair (*e.g.*, of the skin, cornea and mucosal lining). Accordingly, these molecules can be used to treat a variety of cardiovascular diseases including, but not limited to, atherosclerosis, ischemia reperfusion injury, cardiac hypertrophy, hypertension, coronary

artery disease, myocardial infarction, arrhythmia, cardiomyopathies, and congestive heart failure.

As TANGO 272 exhibits expression in the heart, TANGO 272 nucleic acids, proteins, and modulators thereof can be used to treat heart disorders, *e.g.*, ischemic heart disease, atherosclerosis, hypertension, angina pectoris, Hypertrophic Cardiomyopathy, and congenital heart disease.

In another example, TANGO 272 polypeptides, nucleic acids, or modulators thereof, can be used to treat placental disorders, such as toxemia of pregnancy (*e.g.*, preeclampsia and eclampsia), placentitis, or spontaneous abortion.

In another example, TANGO 272 polypeptides, nucleic acids, or modulators thereof, can be used to treat pulmonary (lung) disorders, such as atelectasis, cystic fibrosis, rheumatoid lung disease, pulmonary congestion or edema, chronic obstructive airway disease (*e.g.*, emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis), diffuse interstitial diseases (*e.g.*, sarcoidosis, pneumoconiosis, hypersensitivity pneumonitis, Goodpasture's syndrome, idiopathic pulmonary hemosiderosis, pulmonary alveolar proteinosis, desquamative interstitial pneumonitis, chronic interstitial pneumonia, fibrosing alveolitis, hamman-rich syndrome, pulmonary eosinophilia, diffuse interstitial fibrosis, Wegener's granulomatosis, lymphomatoid granulomatosis, and lipid pneumonia), or tumors (*e.g.*, bronchogenic carcinoma, bronchiolovlveolar carcinoma, bronchial carcinoid, hamartoma, and mesenchymal tumors).

In another example, TANGO 272 polypeptides, nucleic acids, or modulators thereof, can be used to treat hepatic (liver) disorders, such as jaundice, hepatic failure, hereditary hyperbiliruinemias (*e.g.*, Gilbert's syndrome, Crigler-Naijar syndromes and Dubin-Johnson and Rotor's syndromes), hepatic circulatory disorders (*e.g.*, hepatic vein thrombosis and portal vein obstruction and thrombosis) hepatitis (*e.g.*, chronic active hepatitis, acute viral hepatitis, and toxic and drug-induced hepatitis) cirrhosis (*e.g.*, alcoholic cirrhosis, biliary cirrhosis, and hemochromatosis), or malignant tumors (*e.g.*, primary carcinoma, hepatoblastoma, and angiosarcoma).

In another example, TANGO 272 polypeptides, nucleic acids, or modulators thereof, can be used to treat disorders of the brain, such as cerebral edema, hydrocephalus, brain herniations, iatrogenic disease (due to, *e.g.*, infection, toxins, or drugs), inflammations (*e.g.*, bacterial and viral meningitis, encephalitis, and cerebral toxoplasmosis), cerebrovascular diseases (*e.g.*, hypoxia, ischemia, and infarction, intracranial hemorrhage and vascular malformations, and hypertensive encephalopathy), and tumors (*e.g.*, neuroglial tumors, neuronal tumors, tumors of pineal cells, meningeal tumors, primary and secondary lymphomas, intracranial tumors, and medulloblastoma), and to treat injury or trauma to the brain.

In another example, TANGO 272 polypeptides, nucleic acids, or modulators thereof, can be used to treat disorders of skeletal muscle, such as muscular dystrophy (*e.g.*, Duchenne Muscular Dystrophy, Becker Muscular Dystrophy, Emery-Dreifuss Muscular Dystrophy, Limb-Girdle Muscular Dystrophy, Facioscapulohumeral Muscular Dystrophy, Myotonic Dystrophy, Oculopharyngeal Muscular Dystrophy, Distal Muscular Dystrophy, and Congenital Muscular Dystrophy), motor neuron diseases (*e.g.*, Amyotrophic Lateral Sclerosis, Infantile Progressive Spinal Muscular Atrophy, Intermediate Spinal Muscular Atrophy, Spinal Bulbar Muscular Atrophy, and Adult Spinal Muscular Atrophy), myopathies (*e.g.*, inflammatory myopathies (*e.g.*, Dermatomyositis and Polymyositis), Myotonia Congenita, Paramyotonia Congenita, Central Core Disease, Nemaline Myopathy, Myotubular Myopathy, and Periodic Paralysis), and metabolic diseases of muscle (*e.g.*, Phosphorylase Deficiency, Acid Maltase Deficiency, Phosphofructokinase Deficiency, Debrancher Enzyme Deficiency, Mitochondrial Myopathy, Carnitine Deficiency, Carnitine Palmityl Transferase Deficiency, Phosphoglycerate Kinase Deficiency, Phosphoglycerate Mutase Deficiency, Lactate Dehydrogenase Deficiency, and Myoadenylate Deaminase Deficiency).

In another example, TANGO 272 polypeptides, nucleic acids, or modulators thereof, can be used to treat renal disorders, such as glomerular diseases (*e.g.*, acute and chronic glomerulonephritis, rapidly progressive glomerulonephritis, nephrotic syndrome, focal proliferative glomerulonephritis, glomerular lesions associated with systemic disease, such as systemic lupus erythematosus, Goodpasture's syndrome, multiple myeloma, diabetes, neoplasia, sickle cell disease, and chronic inflammatory diseases), tubular diseases (*e.g.*, acute tubular necrosis and acute renal failure, polycystic renal diseasemedullary sponge kidney, medullary cystic disease, nephrogenic diabetes, and renal tubular acidosis), tubulointerstitial diseases (*e.g.*, pyelonephritis, drug and toxin induced tubulointerstitial nephritis, hypercalcemic nephropathy, and hypokalemic nephropathy) acute and rapidly progressive renal failure, chronic renal failure, nephrolithiasis, vascular diseases (*e.g.*, hypertension and nephrosclerosis, microangiopathic hemolytic anemia, atheroembolic renal disease, diffuse cortical necrosis, and renal infarcts), or tumors (*e.g.*, renal cell carcinoma and nephroblastoma).

In another example, TANGO 272 polypeptides, nucleic acids, or modulators thereof, can be used to treat pancreatic disorders, such as pancreatitis (*e.g.*, acute hemorrhagic pancreatitis and chronic pancreatitis), pancreatic cysts (*e.g.*, congenital cysts, pseudocysts, and benign or malignant neoplastic cysts), pancreatic tumors (*e.g.*, pancreatic carcinoma and adenoma), diabetes mellitus (*e.g.*, insulin- and non-insulin-dependent types, impaired glucose tolerance, and gestational diabetes), or islet cell tumors (*e.g.*, insulinomas, adenomas, Zollinger-Ellison syndrome, glucagonomas, and somatostatinoma).

Further, in light of TANGO 272's pattern of expression in humans, TANGO 272 expression can be utilized as a marker for specific tissues (*e.g.*, cardiovascular) and/or cells (*e.g.*, cardiac) in which TANGO 272 is expressed. TANGO 272 nucleic acids can also be utilized for chromosomal mapping.

TANGO 295

A cDNA encoding human TANGO 295 was identified by analyzing the sequences of clones present in a human mammary epithelium cDNA library.

This analysis led to the identification of a clone, jthvb023d09, encoding full-length human TANGO 295. The cDNA of this clone is 1497 nucleotides long (Figure 18; SEQ ID NO:22). The 468 nucleotide open reading frame of this cDNA, nucleotides 217-684 of SEQ ID NO:22 (SEQ ID NO:34), encodes a 156 amino acid protein (Figure 18; SEQ ID NO:23).

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human TANGO 295 includes a 28 amino acid signal peptide at amino acid 1 to about amino acid 28 of SEQ ID NO:23 (SEQ ID NO:125) preceding the mature human TANGO 295 protein which corresponds to about amino acid 29 to amino acid 156 of SEQ ID NO:23 (SEQ ID NO:126).

Human TANGO 295 that has not been post-translationally modified is predicted to have a molecular weight of 17.5 kDa prior to cleavage of its signal peptide and a molecular weight of 14.6 kDa subsequent to cleavage of its signal peptide.

Secretion assays reveal that human TANGO 295 protein is secreted as a 17 kDa protein. The secretion assays were performed as follows: 8×10^5 293T cells were plated per well in a 6-well plate and the cells were incubated in growth medium (DMEM, 10% fetal bovine serum, penicillin/streptomycin) at 37°C, 5% CO_2 overnight. 293T cells were transfected with 2 μg of full-length MANGO 245 inserted in the pMET7 vector/well and 10 μg LipofectAMINE (GIBCO/BRL Cat. # 18324-012) /well according to the protocol for GIBCO/BRL LipofectAMINE. The transfectant was removed 5 hours later and fresh growth medium was added to allow the cells to recover overnight. The medium was removed and each well was gently washed twice with DMEM without methionine and cysteine (ICN Cat. # 16-424-54). 1 ml DMEM without methionine and cysteine with 50 μCi Trans-^{35}S (ICN Cat. # 51006) was added to each well and the cells were incubated at 37°C, 5% CO_2 for the appropriate time period. A 150 μl aliquot of conditioned medium was obtained and 150 μl of 2X SDS sample buffer was added to the aliquot. The sample was heat-inactivated and loaded on a 4-20% SDS-PAGE gel. The gel was fixed and the presence of secreted protein was detected by autoradiography.

Human TANGO 295 includes a pancreatic ribonuclease domain at amino acids 32-156 of SEQ ID NO:23 (SEQ ID NO:97). Figure 20 depicts an alignment of pancreatic ribonuclease domain of human TANGO 295 with a consensus hidden Markov model pancreatic ribonuclease domain (SEQ ID NO:96).

An N-glycosylation site is present at amino acids 127-130 of SEQ ID NO:23. A cAMP/cGMP dependent protein kinase site is present at amino acids 139-142 of SEQ ID NO:23. Protein kinase C phosphorylation sites are present at amino acids 27-29, 62-64, 85-87, and 113-115 of SEQ ID NO:23. N-myristylation sites are present at amino acids 18-23, and 32-37 of SEQ ID NO:23.

Global alignment of the human TANGO 295 and GenPept AF037081 amino acid sequences revealed 53.2% identity (Matrix file used: pam 120.mat, gap penalties of -12/-4; Myers and Miller, 1989, *CABIOS* 4:11-7) (Figure 36). A global alignment of the human TANGO 295 and GenPept AF037081 nucleotide sequences revealed a 22.6% identity between these two sequences (Figures 37A-37C) (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of -2718; Myers and Miller, 1989, *CABIOS* 4:11-7).

Local alignment of the human TANGO 295 and Genbank AF037081 nucleotide sequences revealed 62.7% identity between nucleotides 235-687 of human TANGO 295, and nucleotides 3-453 of AF037081; 43.4% identity between nucleotides 410-850 of human TANGO 295, and nucleotides 3-450 of AF037081; and 46.5% identity between nucleotides 432-700 of human TANGO 295, and nucleotides 5-251 of AF037081 (Matrix file used: pam 120.mat, gap penalties of -12/-4 with a global alignment score of 1214; Huang and Miller, 1991, *Adv. Appl. Math.* 12:373-81) (Figures 38A-38B).

Clone jthvb023d09, which encodes human TANGO 295, was deposited as a composite deposit having a designation EpT295 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on June 18, 1999 and assigned Accession Number PTA-249. Deposit conditions are described below in the section entitled "Deposit of Clones". This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 19 depicts a hydropathy plot of human TANGO 295. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 19 indicates that human TANGO 295 has a signal peptide at its amino terminus, suggesting that human TANGO 295 is a secreted protein.

Use of TANGO 295 Nucleic Acids, Polypeptides, and Modulators Thereof

TANGO 295 includes a pancreatic ribonuclease domain. Proteins having such domains have pyrimidine-specific endonuclease activity, and are present at elevated levels in the pancreas of various mammals and few reptiles. TANGO 295 shows some structural similarities to Ribonuclease k6 (RNase k6). RNase k6 is expressed in human monocytes and monophils (but not in eosinophils), suggesting a role for this ribonuclease in regulating host defense. Based on the structural similarities between TANGO 295 and RNase k6, TANGO 295 may play a role in regulating host defense.

TANGO 295 polypeptides, nucleic acids, and modulators thereof, can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which it is expressed (*e.g.*, mammary epithelium). Accordingly, TANGO 295 polypeptides, nucleic acids, and modulators thereof can be used to treat epithelial disorders, *e.g.*, mammary epithelial disorders (*e.g.*, breast cancer).

Further, in light of TANGO 295's presence in a human mamary epithelium cDNA library, TANGO 295 expression can be utilized as a marker for specific tissues (*e.g.*, breast) and/or cells (*e.g.*,mammary) in which TANGO 295 is expressed. TANGO 295 nucleic acids can also be utilized for chromosomal mapping.

TANGO 354

A cDNA encoding human TANGO 354 was identified by analyzing the sequences of clones present in a Mixed Lymphocyte Reaction (MLR) cDNA library.

This analysis led to the identification of a clone, jthLa042a04, encoding full-length human TANGO 354. The cDNA of this clone is 1788 nucleotides long (Figures 21A-21B; SEQ ID NO:25). The 915 nucleotide open reading frame of this cDNA, nucleotides 62-976 of SEQ ID NO:25 (SEQ ID NO:27), encodes a 305 amino acid protein (Figures 21A-21B; SEQ ID NO:26).

Human TANGO 354 that has not been post-translationally modified is predicted to have a molecular weight of 33.8 kDa prior to cleavage of its signal peptide (31.6 kDa after cleavage of its signal peptide).

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human TANGO 354 includes a 19 amino acid signal peptide at amino acid 1 to about amino acid 19 of SEQ ID NO:26 (SEQ ID NO:127) preceding the mature human TANGO 354 protein which corresponds to about amino acid 20 to amino acid 305 of SEQ ID NO:26 (SEQ ID NO:128).

Human TANGO 354 is a transmembrane protein having an extracellular domain which extends from about amino acid 20 to about amino acid 169 of SEQ ID NO:26 (SEQ ID NO:129), a transmembrane domain which extends from about amino acid 170 to about amino acid 193 of SEQ ID NO:26 (SEQ ID NO:130), and a cytoplasmic domain which extends from about amino acid 194 to amino acid 305 of SEQ ID NO:26 (SEQ ID NO:131).

Alternatively, in another embodiment, a human TANGO 354 protein contains an extracellular domain which extends from about amino acid 194 to amino acid 305 of SEQ ID NO:26 (SEQ ID NO:131), a transmembrane domain which extends from about amino acid 170 to about amino acid 193 of SEQ ID NO:26 (SEQ ID NO:130), and a cytoplasmic domain which extends from about amino acid 20 to about amino acid 169 of SEQ ID NO:26 (SEQ ID NO:129).

Human TANGO 354 includes an immunoglobulin domain at amino acids 33-110 of SEQ ID NO:26 (SEQ ID NO:41). Figure 23 depicts alignments of the immunoglobulin domains of TANGO 354 with consensus hidden Markov model immunoglobulin domains (SEQ ID NO:37).

An N-glycosylation site is present at amino acids 88-91 of SEQ ID NO:26. A cAMP and cGMP-dependent protein kinase phosphorylation site is present at amino acids 233-236 of SEQ ID NO:26. Protein kinase C phosphorylation sites are present at amino acids 81-83, 231-233, and 236-238 of SEQ ID NO:26. Casein kinase II phosphorylation sites are present at amino acids 44-47, 69-72, 81-84, 94-97, 101-104, 113-116, and 146-149 of SEQ ID NO:26. A tyrosine kinase phosphorylation site is present at amino acids 291-299 of SEQ ID NO:26. N-myristylation sites are present at amino acids 30-35, and 109-114 of SEQ ID NO:26.

Clone jthLa042a04, which encodes human TANGO 354, was deposited as EpT354 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on June 18, 1999 and assigned Accession Number PTA-249. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 22 depicts a hydropathy plot of human TANGO 354. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 22 indicates the presence of a hydrophobic domain within human TANGO 354, suggesting that human TANGO 354 is a transmembrane protein.

Use of TANGO 354 Nucleic Acids, Polypeptides, and Modulators Thereof

TANGO 354 includes an immunoglobulin-like domain. Proteins having such domains play a role in mediating protein-protein and protein-ligand interactions, and thus can influence a wide variety of biological processes, including modulation of cell surface recognition; modulation of cellular motility, *e.g.*, chemotaxis and chemokinesis; transduction of an extracellular signal (*e.g.*, by interacting with a ligand and/or a cell-surface receptor); and/or modulation of a signal transduction pathways.

TANGO 354 polypeptides, nucleic acids, and modulators thereof can be used to modulate function, survival, morphology, migration, proliferation and/or differentiation of cells in the tissues in which it is expressed (*e.g.*, hematopoietic tissues).

Because of the presence of an immunoglobulin domain and the expression of TANGO 354 in hematopoietic cells, TANGO 354 polypeptides, nucleic acids, and modulators thereof can be used to modulate (*e.g.*, increase or decrease) hematopoietic function, thereby influencing one or more of: (1) regulation of hematopoiesis; (2) modulation of haemostasis; (3) modulation of an inflammatory response; (4) modulation of neoplastic growth, *e.g.*, inhibition of tumor growth; and/or (5) regulation of thrombolysis.

Accordingly, TANGO 354 polypeptides, nucleic acids, and modulators thereof can be used to treat a variety of hematopoietic diseases including, but not limited to, myeloid disorders and/or lymphoid malignancies. Exemplary myeloid diseases that can be treated include acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus, 1991, *Crit Rev. in Oncol./Hemotol.* 11:267-97). Exemplary lymphoid malignancies that can be treated using these molecules include acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms of malignant lymphomas include non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGF) and Hodgkin's disease.

In one embodiment, TANGO 354 polypeptides, nucleic acids, and modulators thereof can be used to treat a variety of neoplastic diseases, including malignancies of the various organ systems, such as affecting lung, breast, lymphoid, gastrointestinal, and genito-urinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumors, non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus.

The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast carcinomas,

prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and neck, colon and ovary. The term also includes carcinosarcomas, *e.g.*, which include malignant tumors composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures. The term "sarcoma" is art recognized and refers to malignant tumors of mesenchymal derivation.

TANGO 354 polypeptides, nucleic acids, and modulators thereof can also be used to treat a variety of non-cancerous diseases or conditions involving, for example, aberrant T cell activity (*e.g.*, aberrant T cell proliferation and/or secretion). Examples of such T cell diseases or conditions include inflammation; allergy, for example, atopic allergy; organ rejection after transplantation (*e.g.*, skin graft, cardiac graft, islet graft); graft-versus-host disease; autoimmune diseases (including, for example, diabetes mellitus, arthritis (including rheumatoid arthritis, juvenile rheumatoid arthritis, osteoarthritis, psoriatic arthritis), multiple sclerosis, encephalomyelitis, diabetes, myasthenia gravis, systemic lupus erythematosus, autoimmune thyroiditis, dermatitis (including atopic dermatitis and eczematous dermatitis), psoriasis, Sjögren's Syndrome, including keratoconjunctivitis sicca secondary to Sjögren's Syndrome, alopecia areata, allergic responses due to arthropod bite reactions, Crohn's disease, aphthous ulcer, iritis, conjunctivitis, keratoconjunctivitis, ulcerative colitis, asthma, allergic asthma, cutaneous lupus erythematosus, scleroderma, vaginitis, proctitis, drug eruptions, leprosy reversal reactions, erythema nodosum leprosum, autoimmune uveitis, allergic encephalomyelitis, acute necrotizing hemorrhagic encephalopathy, idiopathic bilateral progressive sensorineural hearing loss, aplastic anemia, pure red cell anemia, idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis, chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue, lichen planus, Crohn's disease, Graves ophthalmopathy, sarcoidosis, primary biliary cirrhosis, uveitis posterior, and interstitial lung fibrosis).

Further, in light of TANGO 345's presence in a Mixed Lymphocyte Reaction cDNA library, TANGO 345 expression can be utilized as a marker for specific tissues (*e.g.*, lymphoid tissues such as the thymus and spleen) and/or cells (*e.g.*, lymphocytes) in which TANGO 345 is expressed. TANGO 345 nucleic acids can also be utilized for chromosomal mapping.

TANGO 378

A cDNA encoding human TANGO 378 was identified by analyzing the sequences of clones present in a human natural killer cell cDNA library.

This analysis led to the identification of a clone, jthta028f04, encoding full-length human TANGO 378. The cDNA of this clone is 3258 nucleotides long (Figures 24A-24C; SEQ ID NO:28). The 1584 nucleotide open reading frame of this cDNA, nucleotides 42 to 1625 of SEQ ID NO:28 (SEQ ID NO:30), encodes a 528 amino acid protein (Figure 25; SEQ ID NO:29).

The signal peptide prediction program SIGNALP (Nielsen et al., 1997, *Protein Engineering* 10:1-6) predicted that human TANGO 378 includes a 21 amino acid signal peptide at amino acid 1 to about amino acid 21 of SEQ ID NO:29 (SEQ ID NO:132) preceding the mature human MANGO 347 protein which corresponds to about amino acid 22 to amino acid 528 of SEQ ID NO:29 (SEQ ID NO:133).

Human TANGO 378 that has not been post-translationally modified is predicted to have a molecular weight of 59.0 kDa prior to cleavage of its signal peptide and a molecular weight of 56.7 kDa subsequent to cleavage of its signal peptide.

Human TANGO 378 is a seven transmembrane G-protein coupled receptor (GPCR) protein having an N-terminal extracellular domain which extends from about amino acid 22 to about amino acid 244 of SEQ ID NO:29 (SEQ ID NO:134); seven transmembrane domains which extend from about amino acids 245 to about amino acid 269 of SEQ ID NO:29 (SEQ ID NO:135), about amino acids 287 to about amino acid 306 of SEQ ID NO:29 (SEQ ID NO:136), about amino acids 323 to about amino acid 343 of SEQ ID NO:29 (SEQ ID NO:137), about amino acids 358 to about amino acid 376 of SEQ ID NO:29 (SEQ ID NO:138), about amino acids 414 to about amino acid 438 of SEQ ID NO:29 (SEQ ID NO:139), about amino acids 457 to about amino acid 477 of SEQ ID NO:29 (SEQ ID NO:140), and about amino acids 485 to about amino acid 504 of SEQ ID NO:29 (SEQ ID NO:141); and a C-terminal cytoplasmic domain which extends from about amino acid 505 to amino acid 528 of SEQ ID NO:29 (SEQ ID NO:142). Figure 26 depicts an alignment of each of the transmembrane domains of TANGO 378 with the consensus hidden Markov model seven transmembrane receptor sequences (SEQ ID NO:98).

Alternatively, in another embodiment, a human TANGO 378 protein contains an N-terminal extracellular domain which extends from about amino acid 505 to amino acid 528 of SEQ ID NO:29 (SEQ ID NO:142); seven transmembrane domains which extend from about amino acids 245 to about amino acid 269 of SEQ ID NO:29 (SEQ ID NO:135), about amino acids 287 to about amino acid 306 of SEQ ID NO:29 (SEQ ID NO:136), about amino acids 323 to about amino acid 343 of SEQ ID NO:29 (SEQ ID NO:137), about amino acids 358 to about amino acid 376 of SEQ ID NO:29 (SEQ ID NO:138), about amino acids 414 to about amino acid 438 of SEQ ID NO:29 (SEQ ID NO:139), about amino acids 457 to about amino acid 477 of SEQ ID NO:29 (SEQ ID NO:140), and about amino acids 485 to about amino acid 504 of SEQ ID NO:29 (SEQ ID NO:141); and a C-

terminal cytoplasmic domain which extends from about amino acid 22 to about amino acid 244 of SEQ ID NO:29 (SEQ ID NO:134).

Human TANGO 378 includes three extracellular loops which extend from about amino acid 307 to about amino acid 322 of SEQ ID NO:29 (SEQ ID NO:143), about amino acid 377 to about amino acid 413 of SEQ ID NO:29 (SEQ ID NO:144), and about amino acid 478 to about amino acid 484 of SEQ ID NO:29 (SEQ ID NO:145).

Human TANGO 378 includes three intracellular loops which extend from about amino acid 270 to about amino acid 286 of SEQ ID NO:29 (SEQ ID NO:146), about amino acid 344 to about amino acid 357 of SEQ ID NO:29 (SEQ ID NO:147), and about amino acid 439 to about amino acid 456 of SEQ ID NO:29 (SEQ ID NO:148).

N-glycosylation sites are present at amino acids 18-21, 58-61, 65-68, 146-149, 173-176, 179-182, 394-397, and 400-403 of SEQ ID NO:29. A cAMP and cGMP-dependent protein kinase phosphorylation site is present at amino acids 274-277 of SEQ ID NO:29. Protein kinase C phosphorylation sites are present at amino acids 45-47, 93-95, 375-377, 437-439, 449-451, and 505-507 of SEQ ID NO:29. Casein kinase II phosphorylation sites are present at amino acids 23-26, 29-32, and 510-513 of SEQ ID NO:29. N-myristylation sites are present at amino acids 86-91, 101-106, 157-162, 255-260, 311-316, 420-425, and 467-472 of SEQ ID NO:29. A thiol (cysteine) protease histidine site is present at amino acid 410-420 of SEQ ID NO:29.

Clone jthta028f04, which encodes human TANGO 378, was deposited as EpT378 with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on June 18, 1999 and assigned Accession Number PTA-249. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 25 depicts a hydropathy plot of human TANGO 378. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 25 indicates that human TANGO 378 has a signal peptide at its amino terminus and seven hydrophobic domains within human TANGO 378, suggesting that human TANGO 378 is a transmembrane protein.

Use of TANGO 378 Nucleic Acids, Polypeptides, and Modulators Thereof

TANGO 378 includes a seven transmembrane domain which is typically found in G-protein coupled receptors. Proteins having such a domain play a role in transducing an extracellular signal, *e.g.*, by interacting with a ligand and/or a cell-surface receptor,

followed by mobilization of intracellular molecules that participate in signal transduction pathways (*e.g.*, adenylate cyclase, or phosphatidylinositol 4,5-bisphosphate (PIP_2), inositol 1,4,5-triphosphate (IP_3)).

TANGO 378 polypeptides, nucleic acids, and modulators thereof can be used to modulate function, survival, morphology, migration, proliferation and/or differentiation of cells in the tissues in which it is expressed (*e.g.*, natural killer cells). For example, TANGO 354 polypeptides, nucleic acids, and modulators thereof can be used to modulate an immune response in a subject by, for example, (1) modulating immune cytotoxic responses against pathogenic organisms, *e.g.*, viruses, bacteria, and parasites; (2) by modulating organ rejection after transplantation (*e.g.*, skin graft, cardiac graft, islet graft); (3) by modulating immune recognition and lysis of normal and malignant cells; (4) by modulating T cell diseases; and (5) by controlling neoplastic growth, *e.g.*, inhibition of tumor growth.

Accordingly, TANGO 378 polypeptides, nucleic acids, and modulators thereof can be used to treat a variety of diseases involving aberrant immune responses, for example, aberrant T cell activity (*e.g.*, aberrant T cell proliferation and/or secretion). A non-limiting list of diseases involving aberrant T cell activity is provided in the section entitled "TANGO 354" above.

In other embodiments, TANGO 378 polypeptides, nucleic acids, and modulators thereof can be used to treat a variety of neoplastic diseases, including hematopoietic malignancies and including, but not limited to, myeloid disorders, lymphoid malignancies, and/or malignancies of the various organ systems.). A non-limiting list of such neoplastic diseases is provided in the section entitled "TANGO 354" above.

Further, in light of TANGO 378's presence in a Natral Killer cell cDNA library, TANGO 378 expression can be utilized as a marker for specific tissues (*e.g.*, lymphoid tissues such as the thymus and spleen) and/or cells (*e.g.*, Natural Killer cells) in which TANGO 345 is expressed. TANGO 345 nucleic acids can also be utilized for chromosomal mapping.

Tables 1 and 2 below provide summaries of INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378 sequence information.

TABLE 1: Summary of Sequence Information for INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378

Gene	cDNA	ORF	Polypeptide	Figure	ATCC® Accession Number
INTERCEPT 340 human	SEQ ID NO:1	SEQ ID NO:3	SEQ ID NO:2	Figs. 1A-1B	PTA-250
MANGO 003 human	SEQ ID NO:4	SEQ ID NO:6	SEQ ID NO:5	Figs. 4A-4C	207178
MANGO 003 mouse	SEQ ID NO:7	SEQ ID NO:9	SEQ ID NO:8	Fig. 8	
MANGO 347 human	SEQ ID NO:10	SEQ ID NO:12	SEQ ID NO:11	Fig. 10	PTA-250
TANGO 272 human	SEQ ID NO:13	SEQ ID NO:15	SEQ ID NO:14	Figs. 13A-13D	PTA-250
TANGO 272 mouse	SEQ ID NO:16	SEQ ID NO:18	SEQ ID NO:17	Figs. 16A-16B	
TANGO 272 rat	SEQ ID NO:19	SEQ ID NO:21	SEQ ID NO:20	Figs. 33A-33C	
TANGO 295 human	SEQ ID NO:22	SEQ ID NO:24	SEQ ID NO:23	Fig. 18	PTA-249
TANGO 354 human	SEQ ID NO:25	SEQ ID NO:27	SEQ ID NO:26	Figs. 21A-21B	PTA-249
TANGO 378 human	SEQ ID NO:28	SEQ ID NO:30	SEQ ID NO:29	Figs. 24A-24C	PTA-249

TABLE 2: Summary of Protein Domains of INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378

Protein	Signal Peptide	Mature Protein	Extracellular Domain	Transmembrane Domain	Cytoplasmic Domain
INTERCEPT 340 human	---	---	---	---	---
MANGO 003 human	AA 1-24 of SEQ ID NO:5 SEQ ID NO:101	AA 25-504 of SEQ ID NO:5 SEQ ID NO:102	AA 25-374 of SEQ ID NO:5 SEQ ID NO:103	AA 375-398 of SEQ ID NO:5 SEQ ID NO:104	AA 399-504 of SEQ ID NO:5 SEQ ID NO:105
MANGO 003 mouse	---	AA 1-208 of SEQ ID NO:8 SEQ ID NO:106	AA 1-73 of SEQ ID NO:8 SEQ ID NO:107	AA 74-96 of SEQ ID NO:8 SEQ ID NO:108	AA 97-208 of SEQ ID NO:8 SEQ ID NO:109
MANGO 347 human	AA 1-35 of SEQ ID NO:11 SEQ ID NO:110	AA 36-138 of SEQ IDNO:11 SEQ ID NO:111	---	---	---
TANGO 272 human	AA 1-20 of SEQ ID NO:14 SEQ ID NO:112	AA 21-1050 of SEQ ID NO:14 SEQ ID NO:113	AA 21-767 of SEQ ID NO:14 SEQ ID NO:114	AA 768-791 of SEQ ID NO:14 SEQ ID NO:115	AA 792-1050 of SEQ ID NO:14 SEQ ID NO:116
TANGO 272 mouse	---	AA 1-497 of SEQ ID NO:17 SEQ ID NO:117	AA 1-216 of SEQ ID NO:17 SEQ ID NO:118	AA 217-240 of SEQ ID NO:17 SEQ ID NO:119	AA 241-497 of SEQ ID NO:17 SEQ ID NO:120
TANGO 272 rat	---	AA 1-636 of SEQ ID NO:20 SEQ ID NO:121	AA 1-500 of SEQ ID NO:20 SEQ ID NO:122	AA 501-524 of SEQ ID NO:20 SEQ ID NO:123	AA 525-636 of SEQ ID NO:20 SEQ ID NO:124
TANGO 295 human	AA 1-28 of SEQ ID NO:23 SEQ ID NO:125	AA 29-156 of SEQ ID NO:23 SEQ ID NO:126	---	---	---
TANGO 354 human	AA 1-19 of SEQ ID NO:26 SEQ ID NO:127	AA 20-305 of SEQ ID NO:26 SEQ ID NO:128	AA 20-169 of SEQ ID NO:26 SEQ ID NO:129	AA 170-193 of SEQ ID NO:26 SEQ ID NO:130	AA 194-305 of SEQ ID NO:26 SEQ ID NO:131

TABLE 2 continued

Protein	**Signal Peptide**	**Mature Protein**	**Extracellular Domain**	**Transmembrane Domain**	**Cytoplasmic Domain**
TANGO 378 human	AA 1-21 of SEQ ID NO:29 SEQ ID NO:132	AA 22-528 of SEQ ID NO:29 SEQ ID NO:133	AA 22-244 of SEQ ID NO:29 SEQ ID NO:134	AA 245-269 of SEQ ID NO:29 SEQ ID NO:135 AA 287-306 of SEQ ID NO:29 SEQ ID NO:136 AA 323-343 of SEQ ID NO:29 SEQ ID NO:137 AA 358-376 of SEQ ID NO:29 SEQ ID NO:138 AA 414-438 of SEQ ID NO:29 SEQ ID NO:139 AA 457-477 of SEQ ID NO:29 SEQ ID NO:140 AA 485-504 of SEQ ID NO:29 SEQ ID NO:141	AA 505-528 of SEQ ID NO:29 SEQ ID NO:142

Various aspects of the invention are described in further detail in the following subsections

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode a polypeptide of the invention or a biologically active portion thereof, as well as nucleic acid molecules sufficient for use as hybridization probes to identify nucleic acid molecules encoding a polypeptide of the invention and fragments of such nucleic acid molecules suitable for use as PCR primers for the amplification or mutation of nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid molecule. Preferably, an "isolated" nucleic acid molecule is free of sequences (preferably protein encoding sequences) which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. In other embodiments, the "isolated" nucleic acid is free of intron sequences. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kB, 4 kB, 3 kB, 2 kB, 1 kB, 0.5 kB or 0.1 kB of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. In one embodiment, the nucleic acid molecules of the invention comprise a contiguous open reading frame encoding a polypeptide of the invention.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, or a complement thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequences of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30 as a hybridization probe, nucleic acid molecules of the invention can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook et al., eds., *Molecular Cloning: A Laboratory Manual*, 2nd ed.,1989, Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

A nucleic acid molecule of the invention can be amplified using cDNA, mRNA or genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to all or a portion of a nucleic acid molecule of the invention can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, or a portion thereof. A nucleic acid molecule which is complementary to a given nucleotide

sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence thereby forming a stable duplex.

Moreover, a nucleic acid molecule of the invention can comprise only a portion of a nucleic acid sequence encoding a full length polypeptide of the invention for example, a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a polypeptide of the invention. The nucleotide sequence determined from the cloning one gene allows for the generation of probes and primers designed for use in identifying and/or cloning homologues in other cell types, *e.g.*, from other tissues, as well as homologues from other mammals. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 or 400 consecutive nucleotides of the sense or anti-sense sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, or of a naturally occurring mutant of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30.

Probes based on the sequence of a nucleic acid molecule of the invention can be used to detect transcripts or genomic sequences encoding the same protein molecule encoded by a selected nucleic acid molecule. The probe comprises a label group attached thereto, *e.g.*, a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as part of a diagnostic test kit for identifying cells or tissues which mis-express the protein, such as by measuring levels of a nucleic acid molecule encoding the protein in a sample of cells from a subject, *e.g.*, detecting mRNA levels or determining whether a gene encoding the protein has been mutated or deleted.

A nucleic acid fragment encoding a biologically active portion of a polypeptide of the invention can be prepared by isolating a portion of any of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, expressing the encoded portion of the polypeptide protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the polypeptide.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30.

In addition to the nucleotide sequences of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequence may exist within a population (*e.g.*, the human population). Such genetic polymorphisms may

exist among individuals within a population due to natural allelic variation. An allele is one of a group of genes which occur alternatively at a given genetic locus. As used herein, the phrase “allelic variant” refers to a nucleotide sequence which occurs at a given locus or to a polypeptide encoded by the nucleotide sequence. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a polypeptide of the invention. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a given gene. Alternative alleles can be identified by sequencing the gene of interest in a number of different individuals. This can be readily carried out by using hybridization probes to identify the same genetic locus in a variety of individuals. Any and all such nucleotide variations and resulting amino acid polymorphisms or variations that are the result of natural allelic variation and that do not alter the functional activity are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding proteins of the invention from other species (homologues), which have a nucleotide sequence which differs from that of the human protein described herein are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of a cDNA of the invention can be isolated based on their identity to the human nucleic acid molecule disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a cDNA encoding a soluble form of a membrane-bound protein of the invention isolated based on its hybridization to a nucleic acid molecule encoding all or part of the membrane-bound form. Likewise, a cDNA encoding a membrane-bound form can be isolated based on its hybridization to a nucleic acid molecule encoding all or part of the soluble form.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, or 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000, or 4200) nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence, preferably the coding sequence, of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, or a complement thereof.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% (65%, 70%, preferably 75%) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, 1989, John Wiley & Sons, NY, sections 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are

hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45 C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65 C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, or a complement thereof, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In addition to naturally-occurring allelic variants of a nucleic acid molecule of the invention sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation thereby leading to changes in the amino acid sequence of the encoded protein, without altering the biological activity of the protein. For example, one can make nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are not conserved or only semi-conserved among homologues of various species may be non-essential for activity and thus would be likely targets for alteration. Alternatively, amino acid residues that are conserved among the homologues of various species (*e.g.*, murine and human) may be essential for activity and thus would not be likely targets for alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding a polypeptide of the invention that contain changes in amino acid residues that are not essential for activity. Such polypeptides differ in amino acid sequence from SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule includes a nucleotide sequence encoding a protein that includes an amino acid sequence that is at least about 45% identical, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29.

An isolated nucleic acid molecule encoding a variant protein can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Briefly, PCR primers are designed that delete the trinucleotide codon of the amino acid to be changed and replace it with the trinucleotide codon of the amino acid to be included. This primer is used in the PCR amplification of DNA encoding the protein of interest. This fragment is then isolated and inserted into the full length cDNA

encoding the protein of interest and expressed recombinantly. The resulting protein now includes the amino acid replacement.

Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are can be divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. In similar fashion, the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine histidine, (3) aliphatic = glycine, alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide = asparagine, glutamine; and (6) sulfur - containing = cysteine and methionine. (*See*, for example, Biochemistry, 4th ed., Ed. by L. Stryer, WH Freeman and Co.: 1995).

Alternatively, mutations can be introduced randomly along all or part of the coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for biological activity to identify mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant polypeptide that is a variant of a polypeptide of the invention can be assayed for: (1) the ability to form protein-protein interactions with proteins in a signaling pathway of the polypeptide of the invention; (2) the ability to bind a ligand of the polypeptide of the invention; or (3) the ability to bind to an intracellular target protein of the polypeptide of the invention. In yet another preferred embodiment, the mutant polypeptide can be assayed for the ability to modulate cellular proliferation, cellular migration or chemotaxis, or cellular differentiation.

The present invention encompasses antisense nucleic acid molecules, i.e., molecules which are complementary to a sense nucleic acid encoding a polypeptide of the invention, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire coding strand, or to only a portion thereof, *e.g.*, all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can be antisense to all or part of a non-coding region of the coding strand of a nucleotide sequence encoding a polypeptide of the invention. The non-coding regions ("5' and 3' untranslated regions") are

the 5' and 3' sequences which flank the coding region and are not translated into amino acids.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, β-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, β-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a selected polypeptide of the invention to thereby inhibit expression, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to

receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

An antisense nucleic acid molecule of the invention can be an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier et al., 1987, *Nucleic Acids Res.* 15:6625-41). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al.,1987, *Nucleic Acids Res.* 15:6131-48) or a chimeric RNA-DNA analogue (Inoue et al., 1987, *FEBS Lett.* 215:327-30).

The invention also encompasses ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes; described in Haselhoff and Gerlach, 1988, *Nature* 334:585-91) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of the protein encoded by the mRNA. A ribozyme having specificity for a nucleic acid molecule encoding a polypeptide of the invention can be designed based upon the nucleotide sequence of a cDNA disclosed herein. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, an mRNA encoding a polypeptide of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel and Szostak, 1993, *Science* 261:1411-8.

The invention also encompasses nucleic acid molecules which form triple helical structures. For example, expression of a polypeptide of the invention can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the gene encoding the polypeptide (*e.g.*, the promoter and/or enhancer) to form triple helical structures that prevent transcription of the gene in target cells. *See generally* Helene, 1991, *Anticancer Drug Des.* 6(6):569-84; Helene, 1992, *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, 1992, *Bioassays* 14(12):807-15.

In various embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose

phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (*see* Hyrup et al., 1996, *Bioorganic & Medicinal Chemistry* 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al., 1996, *supra*; Perry-O'Keefe et al., 1996, *Proc. Natl. Acad. Sci. USA* 93:14670-5.

PNAs can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup, 1996, *supra*); or as probes or primers for DNA sequence and hybridization (Hyrup, 1996, *supra*; Perry-O'Keefe et al., 1996, *Proc. Natl. Acad. Sci. USA* 93:14670-675).

In another embodiment, PNAs can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup, 1996, *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996, *supra*) and Finn et al. (1996, *Nucleic Acids Res.* 24(17):3357-63). For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs. Compounds such as 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag et al., 1989, *Nucleic Acids Res.* 17:5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al., 1996, *Nucleic Acids Res.* 24(17):3357-63). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser et al., 1975, *Bioorganic Med. Chem. Lett.* 5:1119-1124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger et al., 1989, *Proc. Natl. Acad. Sci. USA* 86:6553-6; Lemaitre et al., 1987, *Proc. Natl. Acad. Sci. USA* 84:648-52; PCT Publication No. W0 88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. W0 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (*see, e.g.*, Krol et al., 1988, *Bio/Techniques* 6:958-76) or intercalating agents (*see, e.g.*, Zon, 1988, *Pharm. Res.* 5:539-49). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

II. Isolated Proteins and Antibodies

One aspect of the invention pertains to isolated proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise antibodies directed against a polypeptide of the invention. In one embodiment, the native polypeptide can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, polypeptides of the invention are produced by recombinant DNA techniques. Alternative to recombinant expression, a polypeptide of the invention can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, or 5% (by dry weight) of heterologous protein (also referred to herein as a "contaminating protein"). When the protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When the protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly such preparations of the protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or compounds other than the polypeptide of interest. The term "pure" or "isolated" as used herein preferably has the same numerical limits as

"purified" or "isolated" immediately above. "Isolated" and "purified" do not encompass either natural materials in their native state or natural materials that have been separated into components (*e.g.*, in an acrylamide gel) but not obtained either as pure (*e.g.*, lacking contaminating proteins, or chromatography reagents such as denaturing agents and polymers, *e.g.*, acrylamide or agarose) substances or solutions. In preferred embodiments, purified or isolated preparations will lack any contaminating proteins from the same animal from which the protein is normally produced, as can be accomplished by recombinant expression of, for example, a human protein in a non-human cell.

Biologically active portions of a polypeptide of the invention include polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the protein (*e.g.*, the amino acid sequence shown in any of SEQ ID NOs:2, 5, 8, 11, 14, or 17), which include fewer amino acids than the full length protein, and exhibit at least one activity of the corresponding full-length protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the corresponding protein. A biologically active portion of a protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of the native form of a polypeptide of the invention.

Preferred polypeptides have the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29. Other useful proteins are substantially identical (*e.g.*, at least about 45%, preferably 55%, 65%, 75%, 85%, 95%, or 99%) to any of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29 and retain the functional activity of the protein of the corresponding naturally-occurring protein yet differ in amino acid sequence due to natural allelic variation or mutagenesis.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions (*e.g.*, overlapping positions) x 100). In one embodiment the two sequences are the same length.

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990, *Proc. Natl. Acad. Sci. USA* 87:2264-8), modified as in Karlin and Altschul (1993, *Proc. Natl. Acad. Sci. USA* 90:5873-7). Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul et al. (1990, *J. Mol. Biol.* 215:403-10). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to a protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997, *Nucleic Acids Res.* 25:3389-402). Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules (*Id.*). When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. *See* http://www.ncbi.nlm.nih.gov. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller (1988, *CABIOS* 4:11-7). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, typically exact matches are counted.

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises all or part (preferably biologically active) of a polypeptide of the invention operably linked to a heterologous polypeptide (i.e., a polypeptide other than the same polypeptide of the invention). Within the fusion protein, the term "operably linked" is intended to indicate that the polypeptide of the invention and the heterologous polypeptide are fused in-frame to each other. The heterologous polypeptide can be fused to the N-terminus or C-terminus of the polypeptide of the invention.

One useful fusion protein is a GST fusion protein in which the polypeptide of the invention is fused to the C-terminus of GST sequences. Such fusion proteins can facilitate the purification of a recombinant polypeptide of the invention.

In another embodiment, the fusion protein contains a heterologous signal peptide at its N-terminus. For example, the native signal peptide of a polypeptide of the invention can be removed and replaced with a signal peptide from another protein. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal peptide (*Current Protocols in Molecular Biology*, 1992, Ausubel et al., eds., John Wiley & Sons). Other examples of eukaryotic heterologous signal peptides include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal peptides include the phoA secretory signal (Sambrook et al., *supra*) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

In yet another embodiment, the fusion protein is an immunoglobulin fusion protein in which all or part of a polypeptide of the invention is fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand (soluble or membrane-bound) and a protein on the surface of a cell (receptor), to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion protein can be used to affect the bioavailability of a cognate ligand of a polypeptide of the invention. Inhibition of ligand/receptor interaction may be useful therapeutically, both for treating proliferative and differentiative disorders and for modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies directed against a polypeptide of the invention in a subject, to purify ligands and in screening assays to identify molecules which inhibit the interaction of receptors with ligands.

Chimeric and fusion proteins of the invention can be produced by standard recombinant DNA techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel et al., *supra*). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide of the invention.

A signal peptide of a polypeptide of the invention (SEQ ID NOs:101, 110, 112, 125, 127, or 132) can be used to facilitate secretion and isolation of the secreted protein or other proteins of interest. Signal peptides are typically characterized by a core of hydrophobic amino acids which are generally cleaved from the mature protein during secretion in one or

more cleavage events. Such signal peptides contain processing sites that allow cleavage of the signal peptide from the mature proteins as they pass through the secretory pathway. Thus, the invention pertains to the described polypeptides having a signal peptide, as well as to the signal peptide itself and to the polypeptide in the absence of the signal peptide (i.e., the cleavage products). In one embodiment, a nucleic acid sequence encoding a signal peptide of the invention can be operably linked in an expression vector to a protein of interest, such as a protein which is ordinarily not secreted or is otherwise difficult to isolate. The signal peptide directs secretion of the protein, such as from a eukaryotic host into which the expression vector is transformed, and the signal peptide is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by art recognized methods. Alternatively, the signal peptide can be linked to the protein of interest using a sequence which facilitates purification, such as with a GST domain.

In another embodiment, the signal peptides of the present invention can be used to identify regulatory sequences, *e.g.*, promoters, enhancers, repressors. Since signal peptides are the most amino-terminal sequences of a peptide, it is expected that the nucleic acids which flank the signal peptide on its amino-terminal side will be regulatory sequences which affect transcription. Thus, a nucleotide sequence which encodes all or a portion of a signal peptide can be used as a probe to identify and isolate signal peptides and their flanking regions, and these flanking regions can be studied to identify regulatory elements therein.

The present invention also pertains to variants of the polypeptides of the invention. Such variants have an altered amino acid sequence which can function as either agonists (mimetics) or as antagonists. Variants can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation. An agonist can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the protein. An antagonist of a protein can inhibit one or more of the activities of the naturally occurring form of the protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the protein of interest. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein can have fewer side effects in a subject relative to treatment with the naturally occurring form of the protein.

Modification of the structure of the subject polypeptides can be for such purposes as enhancing therapeutic or prophylactic efficacy, stability (*e.g.*, ex vivo shelf life and resistance to proteolytic degradation *in vivo*), or post-translational modifications (*e.g.*, to alter phosphorylation pattern of protein). Such modified peptides, when designed to retain at least one activity of the naturally-occurring form of the protein, or to produce specific

antagonists thereof, are considered functional equivalents of the polypeptides described in more detail herein. Such modified peptides can be produced, for instance, by amino acid substitution, deletion, or addition.

For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (*i.e.* isosteric and/or isoelectric mutations) will not have a major effect on the biological activity of the resulting molecule.

Whether a change in the amino acid sequence of a peptide results in a functional homolog (*e.g.*, functional in the sense that the resulting polypeptide mimics or antagonizes the wild-type form) can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion similar to the wild-type protein, or competitively inhibit such a response. Polypeptides in which more than one replacement has taken place can readily be tested in the same manner.

Variants of a protein of the invention which function as either agonists (mimetics) or as antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the protein of the invention for agonist or antagonist activity. In one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display). There are a variety of methods which can be used to produce libraries of potential variants of the polypeptides of the invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (*see, e.g.,* Narang, 1983, *Tetrahedron* 39:3; Itakura et al., 1984, *Annu. Rev. Biochem.* 53:323; Itakura et al., 1984, *Science* 198:1056; Ike et al., 1983, *Nucleic Acid Res.* 11:477).

In addition, libraries of fragments of the coding sequence of a polypeptide of the invention can be used to generate a variegated population of polypeptides for screening and subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library

can be derived which encodes N-terminal and internal fragments of various sizes of the protein of interest.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify variants of a protein of the invention (Arkin and Yourvan, 1992, *Proc. Natl. Acad. Sci. USA* 89:7811-5; Delgrave et al., 1993, *Protein Engineering* 6(3):327-31).

An isolated polypeptide of the invention, or a fragment thereof, can be used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. The full-length polypeptide or protein can be used or, alternatively, the invention provides antigenic peptide fragments for use as immunogens. The antigenic peptide of a protein of the invention comprises at least 8 (preferably 10, 15, 20, or 30) amino acid residues of the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, and encompasses an epitope of the protein such that an antibody raised against the peptide forms a specific immune complex with the protein.

Preferred epitopes encompassed by the antigenic peptide are regions that are located on the surface of the protein, *e.g.*, hydrophilic regions. Hydropathy plots or similar analyses can be used to identify hydrophilic regions.

An immunogen typically is used to prepare antibodies by immunizing a suitable subject, (*e.g.*, rabbit, goat, mouse or other mammal). An appropriate immunogenic preparation can contain, for example, recombinantly expressed or chemically synthesized polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent.

Accordingly, another aspect of the invention pertains to antibodies directed against a polypeptide of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds an antigen, such as a polypeptide of the invention, *e.g.*, an epitope of a polypeptide of the invention. A molecule which specifically binds to a given polypeptide of the invention is a molecule which binds the polypeptide, but does not substantially bind other molecules in a

sample, *e.g.*, a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a polypeptide of the invention as an immunogen. Preferred polyclonal antibody compositions are ones that have been selected for antibodies directed against a polypeptide or polypeptides of the invention. Particularly preferred polyclonal antibody preparations are ones that contain only antibodies directed against a polypeptide or polypeptides of the invention. Particularly preferred immunogen compositions are those that contain no other human proteins such as, for example, immunogen compositions made using a non-human host cell for recombinant expression of a polypeptide of the invention. In such a manner, the only human epitope or epitopes recognized by the resulting antibody compositions raised against this immunogen will be present as part of a polypeptide or polypeptides of the invention.

The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules can be isolated from the mammal (*e.g.*, from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. Alternatively, antibodies specific for a protein or polypeptide of the invention can be selected for (*e.g.*, partially purified) or purified by, *e.g.*, affinity chromatography. For example, a recombinantly expressed and purified (or partially purified) protein of the invention is produced as described herein, and covalently or non-covalently coupled to a solid support such as, for example, a chromatography column. The column can then be used to affinity purify antibodies specific for the proteins of the invention from a sample containing antibodies directed against a large number of different epitopes, thereby generating a substantially purified antibody composition, *i.e.*, one that is substantially free of contaminating antibodies. By a substantially purified antibody composition is meant, in this context, that the antibody sample contains at most only 30% (by dry weight) of contaminating antibodies directed against epitopes other than those on the desired protein or polypeptide of the invention, and preferably at most 20%, yet more preferably at most 10%, and most preferably at most 5% (by dry weight) of the sample is contaminating antibodies. A purified antibody composition

means that at least 99% of the antibodies in the composition are directed against the desired protein or polypeptide of the invention.

At an appropriate time after immunization, *e.g.*, when the specific antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique (Kohler and Milstein, 1975, *Nature* 256:495-7), the human B cell hybridoma technique (Kozbor et al., 1983, *Immunol. Today* 4:72), the EBV-hybridoma technique (Cole et al., 1985, *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pgs. 77-96) or trioma techniques. The technology for producing hybridomas is well known (*see generally Current Protocols in Immunology*, 1994, Coligan et al.,eds., John Wiley & Sons, Inc., New York, NY). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind the polypeptide of interest, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody directed against a polypeptide of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with the polypeptide of interest. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAPJ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al., 1991, *Bio/Technology* 9:1370-2; Hay et al., 1992, *Hum. Antibod. Hybridomas* 3:81-5; Huse et al., 1989, *Science* 246:1275-81; Griffiths et al., 1993, *EMBO J.* 12:725-34.

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. (See, *e.g.*, Cabilly et al., U.S. Patent No. 4,816,567; and Boss et al., U.S. Patent No. 4,816,397, which are incorporated herein by reference in their entirety.) Humanized antibodies are antibody molecules from non-human species having one or more complementarity determining regions (CDRs) from the non-human species and a framework region from a human immunoglobulin molecule. (See, *e.g.*, Queen, U.S.

Patent No. 5,585,089, which is incorporated herein by reference in its entirety.) Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; U.S. Patent No. 4,816,567; European Patent Application 125,023; Better et al., 1988, *Science* 240:1041-3; Liu et al., 1987, *Proc. Natl. Acad. Sci. USA* 84:3439-43; Liu et al., 1987, *J. Immunol.* 139:3521-6; Sun et al., 1987, *Proc. Natl. Acad. Sci. USA* 84:214-8; Nishimura et al., 1987, *Canc. Res.* 47:999-1005; Wood et al., 1985, *Nature* 314:446-9; and Shaw et al., 1988, *J. Natl. Cancer Inst.* 80:1553-9; Morrison, 1985, *Science* 229:1202-7; Oi et al., 1986, *Bio/Techniques* 4:214; U.S. Patent 5,225,539; Jones et al., 1986, *Nature* 321:522-5; Verhoeyan et al., 1988, *Science* 239:1534; and Beidler et al., 1988, *J. Immunol.* 141:4053-60.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced , for example, using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen, *e.g.*, all or a portion of a polypeptide of the invention. Monoclonal antibodies directed against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, *see, e.g.*, U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806. In addition, companies such as Abgenix, Inc. (Freemont, CA), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

Completely human antibodies which recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, *e.g.*, a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope. (Jespers et al., 1994, *Bio/technology* 12:899-903).

Further, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or

cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxcrubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thiepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (I) (IDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine). The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example , a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, α-interferon, β-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon et al., "Monoclonal Antibodies for Immunotargeting of Drugs in Cancer Therapy," in *Monoclonal Antibodies and Cancer Therapy*, 1985, Reisfeld et al., eds., pgs. 243-56; Hellstrom et al., "Antibodies For Drug Delivery," in *Controlled Drug Delivery 2nd Ed.*, 1987, Robinson et al., eds.; Thorpe, "Antibody Carriers of Cytotoxic Agents in Cancer Therapy: A Review," in *Monoclonal Antibodies '84 Biological and Clinical Applications*, 1985, Pinchera et al., eds, pgs. 475-506; "Analysis, Results, and Future Prospective of the Therapeutic Use of Radiolabeled Antibody in Cancer Therapy," in *Monoclonal Anitbodies for Cancer Detection and Therapy*, 1985, Baldwin et al., eds., pgs. 303-16; and Thorpe et al.,1982, *Immunol. Rev.*, 62:119-58. Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

An antibody directed against a polypeptide of the invention (*e.g.*, monoclonal antibody) can be used to isolate the polypeptide by standard techniques, such as affinity

chromatography or immunoprecipitation. Moreover, such an antibody can be used to detect the protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. The antibodies can also be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, 8-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Further, an antibody (or fragment thereof) can be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, α-interferon, β-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines,

interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon et al., "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies And Cancer Therapy, 1985, Reisfeld et al. (eds.), pgs. 243-56, Alan R. Liss, Inc.; Hellstrom et al., "Antibodies For Drug Delivery", in Controlled Drug Delivery (2nd Ed.), 1987, Robinson et al. (eds.), pgs. 623-53, Marcel Dekker, Inc.; Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in Monoclonal Antibodies '84: Biological And Clinical Applications, 1985, Pinchera et al. (eds.), pgs. 475-506; "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in Monoclonal Antibodies For Cancer Detection And Therapy, 1985, Baldwin et al. (eds.), pgs. 303-16, Academic Press, and Thorpe et al., "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", Immunol. Rev., 1982, 62:119-58.

Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980. Accordingly, in one aspect, the invention provides substantially purified antibodies or fragment thereof, and human or non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29; or an amino acid sequence encoded by the cDNA of a clone deposited as ATCC® Accession Number 207178, ATCC® Accession Number PTA-249, or ATCC® Accession Number PTA-250; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29; an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, or 30, or the cDNA of a clone deposited as ATCC® Accession Number 207178, ATCC® Accession Number PTA-249, or ATCC® Accession Number PTA-250, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. In various embodiments, the substantially purified antibodies of the invention, or fragments thereof, can be human, non-human, chimeric and/or humanized antibodies.

In another aspect, the invention provides human or non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, or an amino acid sequence encoded by the cDNA of a clone deposited as ATCC® Accession Number 207178, ATCC® Accession Number PTA-249, or ATCC® Accession Number PTA-250; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID NOs: 2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID NOs: 2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, or 30, or the cDNA of a clone deposited as ATCC® Accession Number 207178, ATCC® Accession Number PTA-249, or ATCC® Accession Number PTA-250, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. Such non-human antibodies can be goat, mouse, sheep, horse, chicken, rabbit, or rat antibodies. Alternatively, the non-human antibodies of the invention can be chimeric and/or humanized antibodies. In addition, the human or non-human antibodies of the invention can be polyclonal antibodies or monoclonal antibodies.

In still a further aspect, the invention provides monoclonal antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, or an amino acid sequence encoded by the cDNA of a clone deposited as ATCC® Accession Number 207178, ATCC® Accession Number PTA-249, or ATCC® Accession Number PTA-250; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID NOs: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, or 30, or the cDNA of a clone deposited as any of ATCC® Accession Number 207178, ATCC® Accession Number PTA-249, or ATCC® Accession Number PTA-250, or a complement thereof, under conditions of

hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. The monoclonal antibodies can be human, humanized, chimeric and/or non-human antibodies.

The substantially purified antibodies or fragments thereof specifically bind to a signal peptide, a secreted sequence, an extracellular domain, a transmembrane or a cytoplasmic domain cytoplasmic membrane of a polypeptide of the invention. In a particularly preferred embodiment, the substantially purified antibodies or fragments thereof, the human or non-human antibodies or fragments thereof, and/or the monoclonal antibodies or fragments thereof, of the invention specifically bind to a secreted sequence or an extracellular domain of the amino acid sequence of SEQ ID NOs:103, 107, 114, 118, 122, 129, or 134. Preferably, the secreted sequence or extracellular domain to which the antibody, or fragment thereof, binds comprises from about amino acids 25-374 of SEQ ID NO:5 (SEQ ID NO:103), from amino acids 1-73 of SEQ ID NO:8 (SEQ ID NO:107), from amino acids 21-767 of SEQ ID NO:14 (SEQ ID NO:114), from amino acids 1-216 of SEQ ID NO:17 (SEQ ID NO:118), from amino acids 1-500 of SEQ ID NO:20 (SEQ ID NO:122) from amino acids 20-169 of SEQ ID NO:26 (SEQ ID NO:129), and from amino acids 22-244 of SEQ ID NO:29 (SEQ ID NO:134).

Any of the antibodies of the invention can be conjugated to a therapeutic moiety or to a detectable substance. Non-limiting examples of detectable substances that can be conjugated to the antibodies of the invention are an enzyme, a prosthetic group, a fluorescent material, a luminescent material, a bioluminescent material, and a radioactive material.

The invention also provides a kit containing an antibody of the invention conjugated to a detectable substance, and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody of the invention and a pharmaceutically acceptable carrier. In preferred embodiments, the pharmaceutical composition contains an antibody of the invention, a therapeutic moiety, and a pharmaceutically acceptable carrier.

Still another aspect of the invention is a method of making an antibody that specifically recognizes INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378, the method comprising immunizing a mammal with a polypeptide. The polypeptide used as an immunogen comprises an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, or an amino acid sequence encoded by the cDNA of a clone deposited as ATCC® Accession Number 207178, ATCC® Accession Number PTA-249, or ATCC® Accession Number PTA-250; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, an amino acid sequence which is at least 95% identical to the amino acid

sequence of any one of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID NOs: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, or 30, or the cDNA of a clone deposited as ATCC® Accession Number 207178, ATCC® Accession Number PTA-249, or ATCC® Accession Number PTA-250, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. After immunization, a sample is collected from the mammal that contains an antibody that specifically recognizes GPVI. Preferably, the polypeptide is recombinantly produced using a non-human host cell. Optionally, the antibodies can be further purified from the sample using techniques well known to those of skill in the art. The method can further comprise producing a monoclonal antibody-producing cell from the cells of the mammal. Optionally, antibodies are collected from the antibody-producing cell.

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide of the invention (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the

nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, *Gene Expression Technology: Methods in Enzymology*, 1990, Academic Press, San Diego, CA. Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein.

The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic (*e.g.*, *E. coli*) or eukaryotic cells (*e.g.*, insect cells (using baculovirus expression vectors), yeast cells or mammalian cells). Suitable host cells are discussed further in Goeddel, *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988, *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al.,1988, *Gene* 69:301-15) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology*, 1990, Academic Press, San Diego, CA pgs. 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, *Gene Expression Technology: Methods in Enzymology*, 1990, Academic Press, San Diego, CA pgs. 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al., 1992, *Nucleic Acids Res.* 20:2111-8). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari et al., 1987, *EMBO J.* 6:229-34), pMFa (Kurjan and Herskowitz, 1982, *Cell* 30:933-43), pJRY88 (Schultz et al., 1987, *Gene* 54:113-23), pYES2 (Invitrogen Corporation, San Diego, CA), and pPicZ (Invitrogen Corp, San Diego, CA).

Alternatively, the expression vector is a baculovirus expression vector. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith et al., 1983, *Mol. Cell Biol.* 3:2156-65) and the pVL series (Lucklow and Summers, 1989, *Virology* 170:31-9).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987, *Nature* 329:840) and pMT2PC (Kaufman et al., 1987, *EMBO J.* 6:187-95). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook et al., *supra*.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific

regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al., 1987, *Genes Dev.* 1:268-77), lymphoid-specific promoters (Calame and Eaton, 1988, *Adv. Immunol.* 43:235-75), in particular promoters of T cell receptors (Winoto and Baltimore, 1989, *EMBO J.* 8:729-33) and immunoglobulins (Banerji et al., 1983, *Cell* 33:729-40; Queen and Baltimore, 1983, *Cell* 33:741-8), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989, *Proc. Natl. Acad. Sci. USA* 86:5473-7), pancreas-specific promoters (Edlund et al., 1985, *Science* 230:912-6), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss, 1990, *Science* 249:374-9) and the α-fetoprotein promoter (Campes and Tilghman, 1989, *Genes Dev.* 3:537-46).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operably linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to the mRNA encoding a polypeptide of the invention. Regulatory sequences operably linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub et al. (1985, *Reviews - Trends in Genetics* 1(1):22-5).

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic (*e.g.*, *E. coli*) or eukaryotic cell (*e.g.*, insect cells, yeast or mammalian cells).

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and

"transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

In another embodiment, the expression characteristics of an endogenous (*e.g.*, INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378) nucleic acid within a cell, cell line or microorganism may be modified by inserting a DNA regulatory element heterologous to the endogenous gene of interest into the genome of a cell, stable cell line or cloned microorganism such that the inserted regulatory element is operatively linked with the endogenous gene (*e.g.*, INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378) and controls, modulates or activates the endogenous gene. For example, endogenous INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378 which are normally "transcriptionally silent", *i.e.*, INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378 genes which are normally not expressed, or are expressed only at very low levels in a cell line or microorganism, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell line or microorganism. Alternatively, transcriptionally silent, endogenous INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378 genes may be activated by insertion of a promiscuous regulatory element that works across cell types.

A heterologous regulatory element may be inserted into a stable cell line or cloned microorganism, such that it is operatively linked with and activates expression of endogenous INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378 genes, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art, and described *e.g.*, in

Chappel, U.S. Patent No. 5,272,071; PCT publication No. WO 91/06667, published May 16, 1991.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce a polypeptide of the invention. Accordingly, the invention further provides methods for producing a polypeptide of the invention using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which a sequences encoding a polypeptide of the invention have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous sequences encoding a polypeptide of the invention have been introduced into their genome or homologous recombinant animals in which endogenous encoding a polypeptide of the invention sequences have been altered. Such animals are useful for studying the function and/or activity of the polypeptide and for identifying and/or evaluating modulators of polypeptide activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing nucleic acid encoding a polypeptide of the invention (or a homologue thereof) into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the polypeptide of the invention to particular cells. Methods for

generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent NOs. 4,736,866; 4,870,009; 4,873,191 and in Hogan (*Manipulating the Mouse Embryo*, 1986, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of mRNA encoding the transgene in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying the transgene can further be bred to other transgenic animals carrying other transgenes.

To create an homologous recombinant animal, a vector is prepared which contains at least a portion of a gene encoding a polypeptide of the invention into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the gene. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous protein). In the homologous recombination vector, the altered portion of the gene is flanked at its 5' and 3' ends by additional nucleic acid of the gene to allow for homologous recombination to occur between the exogenous gene carried by the vector and an endogenous gene in an embryonic stem cell. The additional flanking nucleic acid sequences are of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (*see, e.g.*, Thomas and Capecchi, 1987, *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous gene are selected (*see, e.g.*, Li et al., 1992, *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras (*see, e.g.*, Bradley in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, 1987, Robertson, ed., IRL, Oxford pgs. 113-52). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in

Bradley, 1991, *Current Opinion in Bio/Technology* 2:823-9 and in PCT Publication NOs. WO 90/11354, WO 91/01140, WO 92/0968 and WO 93/04169.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, *see, e.g.*, Lakso et al., 1992, *Proc. Natl. Acad. Sci. USA* 89:6232-6. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al., 1991, *Science* 251:1351-5). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut et al., 1997, *Nature* 385:810-3 and PCT Publication NOs. WO 97/07668 and WO 97/07669.

IV. Pharmaceutical Compositions

The nucleic acid molecules, polypeptides, and antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

The invention includes methods for preparing pharmaceutical compositions for modulating the expression or activity of a polypeptide or nucleic acid of the invention. Such methods comprise formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention. Such compositions can further include additional active agents. Thus, the invention further includes methods for preparing a pharmaceutical composition by formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention and one or more additional active compounds.

The agent which modulates expression or activity may, for example, be a small molecule. For example, such small molecules include peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds. It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention. Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (*e.g.* about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols,

glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor ELJ (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a polypeptide or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral

therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed.

Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

For antibodies, the preferred dosage is 0.1 mg/kg to 100 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (*e.g.*, into the brain). A method for lipidation of antibodies is described by Cruikshank et al. (1997, *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight.

The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody can include a single treatment or, preferably, can include a series of treatments. In a preferred example, a subject is treated with antibody, protein, or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody, protein, or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Patent 5,328,470) or by stereotactic injection (*see, e.g.*, Chen et al., 1994, *Proc. Natl. Acad. Sci. USA* 91:3054-7). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.* retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. <u>Uses and Methods of the Invention</u>

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) detection assays (*e.g.*, chromosomal mapping, tissue typing, forensic biology); c) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and d) methods of treatment (*e.g.*, therapeutic and prophylactic). For example, the INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, and TANGO 378 polypeptides of the invention can to used to modulate cellular function, survival, morphology, proliferation, and/or differentiation of the cells in which they are expressed. For example, the polypeptides of the invention can be used to treat diseases such as neoplastic disorders (*e.g.*, cancer, tumors), hematopoietic disorders (*e.g.*, T cell disorders), among others. The isolated nucleic acid molecules of the invention can be used to express proteins (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect mRNA (*e.g.*, in a biological sample) or a genetic lesion, and to modulate activity of a polypeptide of the invention. In addition, the polypeptides of the invention can be used to screen drugs or compounds which modulate activity or expression of a polypeptide of the invention as well as to treat disorders characterized by insufficient or excessive production of a protein of the invention or production of a form of a protein of the invention which has decreased or aberrant activity compared to the wild type protein. In addition, the antibodies of the invention can be used to detect and isolate a protein of the invention and modulate activity of a protein of the invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

A. Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) which bind to polypeptide of the invention or have a stimulatory or inhibitory effect on, for example, expression or activity of a polypeptide of the invention.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a polypeptide of the invention or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, 1997, *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al., 1993, *Proc. Natl. Acad. Sci. USA* 90:6909; Erb et al., 1994, *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann et al., 1994, *J. Med. Chem.* 37:2678; Cho et al., 1993, *Science* 261:1303; Carrell et al., 1994, *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell et al., 1994, *Angew. Chem. Int. Ed. Engl.* 33:2061; and Gallop et al., 1994, *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992, *Bio/Techniques* 13:412-21), or on beads (Lam, 1991, *Nature* 354:82-4), chips (Fodor, 1993, *Nature* 364:555-6), bacteria (U.S. Patent No. 5,223,409), spores (U.S. Patent NOs. 5,571,698; 5,403,484; and 5,223,409), plasmids (Cull et al., 1992, *Proc. Natl. Acad. Sci. USA* 89:1865-9) or phage (Scott and Smith, 1990, *Science* 249:386-90; Devlin, 1990, *Science* 249:404-6; Cwirla et al., 1990, *Proc. Natl. Acad. Sci. USA* 87:6378-82; and Felici, 1991, *J. Mol. Biol.* 222:301-10).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to the polypeptide determined. The cell, for example, can be a yeast cell or a cell of mammalian origin. Determining the ability of the test compound to bind to the polypeptide can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the polypeptide or

biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In a preferred embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or a biologically active portion thereof as compared to the known compound.

In another embodiment, the assay involves assessment of an activity characteristic of the polypeptide, wherein binding of the test compound with the polypeptide or a biologically active portion thereof alters (*e.g.*, increases or decreases) the activity of the polypeptide.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the polypeptide or a biologically active portion thereof can be accomplished, for example, by determining the ability of the polypeptide protein to bind to or interact with a target molecule or to transport molecules across the cytoplasmic membrane.

Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by one of the methods described above for determining direct binding. As used herein, a "target molecule" is a molecule with which a selected polypeptide (*e.g.*, a polypeptide of the invention binds or interacts with in nature, for example, a molecule on the surface of a cell which expresses the selected protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A target molecule can be a polypeptide of the invention or some other polypeptide or protein. For example, a target molecule can be a component of a signal transduction pathway which facilitates transduction of an extracellular signal (*e.g.*, a signal generated by binding of a

compound to a polypeptide of the invention) through the cell membrane and into the cell or a second intercellular protein which has catalytic activity or a protein which facilitates the association of downstream signaling molecules with a polypeptide of the invention. Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*e.g.*, intracellular Ca^{2+}, diacylglycerol, IP3, etc.), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (*e.g.*, a regulatory element that is responsive to a polypeptide of the invention operably linked to a nucleic acid encoding a detectable marker, *e.g.* luciferase), or detecting a cellular response, for example, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to the polypeptide or biologically active portion thereof. Binding of the test compound to the polypeptide can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished, for example, by determining the ability of the polypeptide to bind to a target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished by determining the ability of the polypeptide of the invention to further modulate the target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay comprises contacting a polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the polypeptide to preferentially bind to or modulate the activity of a target molecule.

The cell-free assays of the present invention are amenable to use of both a soluble form or the membrane-bound form of a polypeptide of the invention. In the case of cell-free assays comprising the membrane-bound form of the polypeptide, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the polypeptide is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-octylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton X-100, Triton X-114, Thesit, Isotridecypoly(ethylene glycol ether)n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either the polypeptide of the invention or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to the polypeptide, or interaction of the polypeptide with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase fusion proteins or glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical; St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or A polypeptide of the invention, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components and complex formation is measured either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of binding or activity of the polypeptide of the invention can be determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the polypeptide of the invention or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated polypeptide of the invention or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (*e.g.*, biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the polypeptide of the invention or target molecules but which do not interfere with binding of the polypeptide of the invention to its target molecule can be derivatized to the wells of the plate, and unbound target or polypeptide of the invention trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the polypeptide of the invention or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the polypeptide of the invention or target molecule.

In another embodiment, modulators of expression of a polypeptide of the invention are identified in a method in which a cell is contacted with a candidate compound and the expression of the selected mRNA or protein (i.e., the mRNA or protein corresponding to a polypeptide or nucleic acid of the invention) in the cell is determined. The level of expression of the selected mRNA or protein in the presence of the candidate compound is compared to the level of expression of the selected mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of expression of the polypeptide of the invention based on this comparison. For example, when expression of the selected mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of the selected mRNA or protein expression. Alternatively, when expression of the selected mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of the selected mRNA or protein expression. The level of the selected mRNA or protein expression in the cells can be determined by methods described herein.

In yet another aspect of the invention, a polypeptide of the inventions can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos et al., 1993, *Cell* 72:223-32; Madura et al., 1993, *J. Biol. Chem.* 268:12046-54; Bartel et al., 1993, *Bio/Techniques* 14:920-4; Iwabuchi et al., 1993, *Oncogene* 8:1693-6; and PCT Publication No. WO 94/10300), to identify other proteins, which bind to or interact with the polypeptide of the invention and modulate activity of the

polypeptide of the invention. Such binding proteins are also likely to be involved in the propagation of signals by the polypeptide of the inventions as, for example, upstream or downstream elements of a signaling pathway involving the polypeptide of the invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. Accordingly, nucleic acid molecules described herein or fragments thereof, can be used to map the location of the corresponding genes on a chromosome. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the sequence of a gene of the invention. Computer analysis of the sequence of a gene of the invention can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the gene sequences will yield an amplified fragment. For a review of this technique, see D'Eustachio et al. (1983, *Science* 220:919-24).

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the nucleic acid sequences of the invention to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a gene to its chromosome include *in situ* hybridization (described in Fan et al., 1990, *Proc. Natl. Acad. Sci. USA* 87:6223-7), pre-screening with labeled flow-sorted chromosomes (CITE),

and pre-selection by hybridization to chromosome specific cDNA libraries. Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. For a review of this technique, see Verma et al., *Human Chromosomes: A Manual of Basic Techniques*, 1988, Pergamon Press, NY.

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland et al., 1987, *Nature* 325:783-7.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a gene of the invention can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Furthermore, the nucleic acid sequences disclosed herein can be used to perform searches against "mapping databases", *e.g.*, BLAST-type search, such that the chromosome position of the gene is identified by sequence homology or identity with known sequence fragments which have been mapped to chromosomes.

A polypeptide and fragments and sequences thereof and antibodies specific thereto can be used to map the location of the gene encoding the polypeptide on a chromosome. This mapping can be carried out by specifically detecting the presence of the polypeptide in members of a panel of somatic cell hybrids between cells of a first species of animal from which the protein originates and cells from a second species of animal and then determining which somatic cell hybrid(s) expresses the polypeptide and noting the chromosome(s) from

the first species of animal that it contains. For examples of this technique, see Pajunen et al., 1988, *Cytogenet. Cell Genet.* 47:37-41 and Van Keuren et al., 1986, *Hum. Genet.* 74:34-40. Alternatively, the presence of the polypeptide in the somatic cell hybrids can be determined by assaying an activity or property of the polypeptide, for example, enzymatic activity, as described in Bordelon-Riser et al., 1979, *Somatic Cell Genetics* 5:597-613 and Owerbach et al., 1978, *Proc. Natl. Acad. Sci. USA* 75:5640-5644.

2. Tissue Typing

The nucleic acid sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the nucleic acid sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The nucleic acid sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NOs:1, 4, 7, 10, 13, 16, 19, 22, 25, and 28 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NOs:3, 6, 9, 12, 15, 18, 21, 24, 27,

and 30 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from the nucleic acid sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial Gene Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the nucleic acid sequences of the invention or portions thereof, *e.g.*, fragments derived from noncoding regions having a length of at least 20 or 30 bases.

The nucleic acid sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such probes can be used to identify tissue by species and/or by organ type.

C. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic

(predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, or TANGO 378 protein and/or nucleic acid expression as well as INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, or TANGO 378 activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant or unwanted INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, or TANGO 378 gene expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, or TANGO 378 protein or nucleic acid expression or activity. For example, mutations in a gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with protein or nucleic acid expression or activity.

As an alternative to making determinations based on the absolute expression level of selected genes, determinations may be based on the normalized expression levels of these genes. Expression levels are normalized by correcting the absolute expression level of a INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, or TANGO 378 gene by comparing its expression to the expression of a gene that is not a INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, or TANGO 378, *e.g.*, a housekeeping gene that is constitutively expressed. Suitable genes for normalization include housekeeping genes such as the actin gene. This normalization allows the comparison of the expression level in one sample, *e.g.*, a patient sample, to another sample, *e.g.*, a non-disease sample, or between samples from different sources.

Alternatively, the expression level can be provided as a relative expression level. To determine a relative expression level of a gene, the level of expression of the gene is determined for 10 or more samples of different cell isolates, preferably 50 or more samples, prior to the determination of the expression level for the sample in question. The mean expression level of each of the genes assayed in the larger number of samples is determined and this is used as a baseline expression level for the gene(s) in question. The expression level of the gene determined for the test sample (absolute level of expression) is then divided by the mean expression value obtained for that gene. This provides a relative expression level and aids in identifying extreme cases of disease.

Preferably, the samples used in the baseline determination will be from diseased or from non-diseased cells of tissue. The choice of the cell source is dependent on the use of

the relative expression level. Using expression found in normal tissues as a mean expression score aids in validating whether the INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, or TANGO 378 gene assayed is diseased cell-type specific (versus normal cells). Such a use is particularly important in identifying whether a INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, or TANGO 378 gene can serve as a target gene. In addition, as more data is accumulated, the mean expression value can be revised, providing improved relative expression values based on accumulated data. Expression data from cells provide a means for grading the severity of the disease state.

Another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of INTERCEPT 340, MANGO 003, MANGO 347, TANGO 272, TANGO 295, TANGO 354, or TANGO 378 genes in clinical trials.

These and other agents are described in further detail in the following sections.

1. <u>Diagnostic Assays</u>

An exemplary method for detecting the presence or absence of a polypeptide or nucleic acid of the invention in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting a polypeptide or nucleic acid (*e.g.*, mRNA, genomic DNA) of the invention such that the presence of a polypeptide or nucleic acid of the invention is detected in the biological sample. A preferred agent for detecting mRNA or genomic DNA encoding a polypeptide of the invention is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA encoding a polypeptide of the invention. The nucleic acid probe can be, for example, a full-length cDNA, such as the nucleic acid of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28 or 30, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a mRNA or genomic DNA encoding a polypeptide of the invention. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting a polypeptide of the invention is an antibody capable of binding to a polypeptide of the invention, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly

labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of a polypeptide of the invention include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of a polypeptide of the invention include introducing into a subject a labeled antibody directed against the polypeptide. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting a polypeptide of the invention or mRNA or genomic DNA encoding a polypeptide of the invention, such that the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide is detected in the biological sample, and comparing the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the control sample with the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the test sample.

The invention also encompasses kits for detecting the presence of a polypeptide or nucleic acid of the invention in a biological sample (a test sample). Such kits can be used to determine if a subject is suffering from or is at increased risk of developing a disorder associated with aberrant expression of a polypeptide of the invention (*e.g.*, a proliferative disorder, *e.g.*, psoriasis or cancer). For example, the kit can comprise a labeled compound or agent capable of detecting the polypeptide or mRNA encoding the polypeptide in a biological sample and means for determining the amount of the polypeptide or mRNA in the sample (*e.g.*, an antibody which binds the polypeptide or an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide). Kits can also include instructions for observing that the tested subject is suffering from or is at risk of developing

a disorder associated with aberrant expression of the polypeptide if the amount of the polypeptide or mRNA encoding the polypeptide is above or below a normal level.

For antibody-based kits, the kit can comprise, for example: (1) a first antibody (*e.g.*, attached to a solid support) which binds to a polypeptide of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

For oligonucleotide-based kits, the kit can comprise, for example: (1) an oligonucleotide, *e.g.*, a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule encoding a polypeptide of the invention. The kit can also comprise, *e.g.*, a buffering agent, a preservative, or a protein stabilizing agent. The kit can also comprise components necessary for detecting the detectable agent (*e.g.*, an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit is usually enclosed within an individual container and all of the various containers are within a single package along with instructions for observing whether the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide.

2. <u>Prognostic Assays</u>

The methods described herein can furthermore be utilized as diagnostic or prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with aberrant expression or activity of a polypeptide of the invention. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing such a disease or disorder. Thus, the present invention provides a method in which a test sample is obtained from a subject and a polypeptide or nucleic acid (*e.g.*, mRNA, genomic DNA) of the invention is detected, wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the polypeptide. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to

treat a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, such methods can be used to determine whether a subject can be effectively treated with a specific agent or class of agents (*e.g.*, agents of a type which decrease activity of the polypeptide). Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant expression or activity of a polypeptide of the invention in which a test sample is obtained and the polypeptide or nucleic acid encoding the polypeptide is detected (*e.g.*, wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant expression or activity of the polypeptide).

The methods of the invention can also be used to detect genetic lesions or mutations in a gene of the invention, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized aberrant expression or activity of a polypeptide of the invention. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion or mutation characterized by at least one of an alteration affecting the integrity of a gene encoding the polypeptide of the invention, or the mis-expression of the gene encoding the polypeptide of the invention. For example, such genetic lesions or mutations can be detected by ascertaining the existence of at least one of: 1) a deletion of one or more nucleotides from the gene; 2) an addition of one or more nucleotides to the gene; 3) a substitution of one or more nucleotides of the gene; 4) a chromosomal rearrangement of the gene; 5) an alteration in the level of a messenger RNA transcript of the gene; 6) an aberrant modification of the gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; 8) a non-wild type level of a the protein encoded by the gene; 9) an allelic loss of the gene; and 10) an inappropriate post-translational modification of the protein encoded by the gene. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a gene.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent NOs. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran et al., 1988, *Science* 241:1077-80; and Nakazawa et al., 1994, *Proc. Natl. Acad. Sci. USA* 91:360-4), the latter of which can be particularly useful for detecting point mutations in a gene (*see, e.g.*, Abravaya et al., 1995, *Nucleic Acids Res.* 23:675-82). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to the selected

gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli et al., 1990, *Proc. Natl. Acad. Sci. USA* 87:1874-78), transcriptional amplification system (Kwoh, et al., 1989, *Proc. Natl. Acad. Sci. USA* 86:1173-7), Q-Beta Replicase (Lizardi et al., 1988, *Bio/Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a selected gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin et al., 1996, *Human Mutation* 7:244-55; Kozal et al., 1996, *Nature Medicine* 2:753-9). For example, genetic mutations can be identified in two-dimensional arrays containing light-generated DNA probes as described in Cronin et al., *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the selected gene and detect mutations by comparing the sequence of the sample nucleic acids with the corresponding wild-type (control)

sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert (1977, *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger (1977, *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays developed by Naeve et al. (1995, *Bio/Techniques* 19:448-53), including sequencing by mass spectrometry (*see, e.g.*, PCT Publication No. WO 94/16101; Cohen et al., 1996, *Adv. Chromatogr.* 36:127-62; and Griffin et al., 1993, *Appl. Biochem. Biotechnol.* 38:147-59).

Other methods for detecting mutations in a selected gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al., 1985, *Science* 230:1242). In general, the technique of mismatch cleavage entails providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. RNA/DNA duplexes can be treated with RNase to digest mismatched regions, and DNA/DNA hybrids can be treated with S1 nuclease to digest mismatched regions.

In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.*, Cotton et al., 1988, *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba et al., 1992, *Methods Enzymol.* 217:286-95. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called DNA mismatch repair enzymes) in defined systems for detecting and mapping point mutations in cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al., 1994, *Carcinogenesis* 15:1657-62). According to an exemplary embodiment, a probe based on a selected sequence, *e.g.*, a wild-type sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.*, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type

nucleic acids (Orita et al., 1989, *Proc. Natl. Acad. Sci. USA* 86:2766; *see also* Cotton, 1993, *Mutat. Res.* 285:125-44; Hayashi, 1992, *Genet. Anal. Tech. Appl.* 9:73-9). Single-stranded DNA fragments of sample and control nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, and the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al., 1991, *Trends Genet.* 7:5).

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al., 1985, *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a `GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner, 1987, *Biophys. Chem.* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al., 1986, *Nature* 324:163; Saiki et al., 1989, *Proc. Natl. Acad. Sci. USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; Gibbs et al., 1989, *Nucleic Acids Res.* 17:2437-48) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent or reduce polymerase extension (Prossner, 1993, *Tibtech* 11:238). In addition, it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al., 1992, *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be

performed using Taq ligase for amplification (Barany, 1991, *Proc. Natl. Acad. Sci. USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a gene encoding a polypeptide of the invention. Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which the polypeptide of the invention is expressed may be utilized in the prognostic assays described herein.

3. Pharmacogenomics

Agents, or modulators which have a stimulatory or inhibitory effect on activity or expression of a polypeptide of the invention as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders associated with aberrant activity of the polypeptide. In conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of a polypeptide of the invention, expression of a nucleic acid of the invention, or mutation content of a gene of the invention in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. *See*, *e.g.*, Linder, 1997, *Clin. Chem.* 43(2):254-66. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body are referred to as "altered drug action." Genetic conditions transmitted as single factors altering the way the body acts on drugs are referred to as "altered drug metabolism". These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main

clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, a PM will show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of a polypeptide of the invention, expression of a nucleic acid encoding the polypeptide, or mutation content of a gene encoding the polypeptide in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a modulator of activity or expression of the polypeptide, such as a modulator identified by one of the exemplary screening assays described herein.

4. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of a polypeptide of the invention (*e.g.*, the ability to modulate aberrant cell proliferation chemotaxis, and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent, as determined by a screening assay as described herein, to increase gene expression, protein levels or protein activity, can be monitored in clinical trials of subjects exhibiting decreased

gene expression, protein levels, or protein activity. Alternatively, the effectiveness of an agent, as determined by a screening assay, to decrease gene expression, protein levels or protein activity, can be monitored in clinical trials of subjects exhibiting increased gene expression, protein levels, or protein activity. In such clinical trials, expression or activity of a polypeptide of the invention and preferably, that of other polypeptide that have been implicated in for example, a cellular proliferation disorder, can be used as a marker of the immune responsiveness of a particular cell.

For example, and not by way of limitation, genes, including those of the invention, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) which modulates activity or expression of a polypeptide of the invention (*e.g.*, as identified in a screening assay described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of a gene of the invention and other genes implicated in the disorder. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of a gene of the invention or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of the polypeptide or nucleic acid of the invention in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level the of the polypeptide or nucleic acid of the invention in the post-administration samples; (v) comparing the level of the polypeptide or nucleic acid of the invention in the pre-administration sample with the level of the polypeptide or nucleic acid of the invention in the post-administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of the polypeptide to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of the polypeptide to lower levels than detected, i.e., to decrease the effectiveness of the agent.

C. Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant expression or activity of a polypeptide of the invention, *e.g.*, cardiac infection (*e.g.*, myocarditis or dilated cardiomyopathy), central nervous system infection (*e.g.*, non-specific febrile illness or meningoencephalitis), pancreatic infection (*e.g.*, acute pancreatitis), respiratory infection (pneumonia), gastrointestinal infection, type I diabetes, cancer, familia hypercholesterolemia, treat hemophilia B, Marfan syndrome, protein S deficiency, allergy, inflammation, and gastroduodenal ulcer. Moreover, the polypeptides of the invention can be used to modulate cellular function, survival, morphology, proliferation and/or differentiation.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant expression or activity of a polypeptide of the invention, by administering to the subject an agent which modulates expression or at least one activity of the polypeptide. Subjects at risk for a disease which is caused or contributed to by aberrant expression or activity of a polypeptide of the invention can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of aberrancy, for example, an agonist or antagonist agent can be used for treating the subject.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating expression or activity of a polypeptide of the invention for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of the polypeptide. An agent that modulates activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of the polypeptide, a peptide, a peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more of the biological activities of the polypeptide. Examples of such stimulatory agents include the active polypeptide of the invention and a nucleic acid molecule encoding the polypeptide of the invention that has been introduced into the cell. In another embodiment, the agent inhibits one or more of the biological activities of the polypeptide of the invention. Examples of such inhibitory agents include antisense nucleic acid molecules and antibodies. These modulatory methods can be performed *in vitro* (*e.g.*,

by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a polypeptide of the invention. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, upregulates or downregulates) expression or activity. In another embodiment, the method involves administering a polypeptide of the invention or a nucleic acid molecule of the invention as therapy to compensate for reduced or aberrant expression or activity of the polypeptide.

Stimulation of activity is desirable in situations in which activity or expression is abnormally low or downregulated and/or in which increased activity is likely to have a beneficial effect. Conversely, inhibition of activity is desirable in situations in which activity or expression is abnormally high or upregulated and/or in which decreased activity is likely to have a beneficial effect.

The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

Deposit of Clones

Clones containing cDNA molecules encoding human MANGO 003 were deposited with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on March 30, 1999 as Accession Number 207178, as part of a composite deposit representing a mixture of three strains, each carrying one recombinant plasmid harboring a particular cDNA clone.

To distinguish the strains and isolate a strain harboring a particular cDNA clone, an aliquot of the mixture can be streaked out to single colonies on nutrient medium (*e.g.*, LB plates) supplemented with 100 g/ml ampicillin, single colonies grown, and then plasmid DNA extracted using a standard minipreparation procedure. Next, a sample of the DNA minipreparation can be digested with a combination of the restriction enzymes *Sal* I and *Not* I, and the resultant products resolved on a 0.8% agarose gel using standard DNA electrophoresis conditions. The digest liberates fragments as follows:

human MANGO 003 (clone EpthLa6a1): 3.2 kB

The identity of the strains can be inferred from the fragments liberated.

Clones containing cDNA molecules encoding human INTERCEPT 340, MANGO 347, and TANGO 272 were deposited with the American Type Culture Collection (ATCC®

10801 University Boulevard, Manassas, VA 20110-2209) on June 18, 1999 as Accession Number PTA-250, as part of a composite deposit representing a mixture of three strains, each carrying one recombinant plasmid harboring a particular cDNA clone.

To distinguish the strains and isolate a strain harboring a particular cDNA clone, an aliquot of the mixture can be streaked out to single colonies on nutrient medium (*e.g.*, LB plates) supplemented with 100 g/ml ampicillin, single colonies grown, and then plasmid DNA extracted using a standard minipreparation procedure. Next, a sample of the DNA minipreparation can be digested with a combination of the restriction enzymes *Sal* I and *Not* I, and the resultant products resolved on a 0.8% agarose gel using standard DNA electrophoresis conditions. The digest liberates fragments as follows:

human INTERCEPT 340 (clone EpI340): 3.3 kB
human MANGO 347 (clone EpM347): 1.4 kB
human TANGO 272 (clone EpT272): 5.0 kB

The identity of the strains can be inferred from the fragments liberated.

Clones containing cDNA molecules encoding human TANGO 295, TANGO 354, and TANGO 378 were deposited with the American Type Culture Collection (ATCC® 10801 University Boulevard, Manassas, VA 20110-2209) on June 18, 1999 as Accession Number PTA-249, as part of a composite deposit representing a mixture of three strains, each carrying one recombinant plasmid harboring a particular cDNA clone.

To distinguish the strains and isolate a strain harboring a particular cDNA clone, an aliquot of the mixture can be streaked out to single colonies on nutrient medium (*e.g.*, LB plates) supplemented with 100 g/ml ampicillin, single colonies grown, and then plasmid DNA extracted using a standard minipreparation procedure. Next, a sample of the DNA minipreparation can be digested with a combination of the restriction enzymes *Sal* I and *Not* I, and the resultant products resolved on a 0.8% agarose gel using standard DNA electrophoresis conditions. The digest liberates fragments as follows:

human TANGO 295 (clone EpT295): 1.5 kB
human TANGO 354 (clone EpT354): 1.8 kB
human TANGO 378 (clone EpT378): 3.3 kB

The identity of the strains can be inferred from the fragments liberated.

All publications, patents and patent applications mentioned in this specification are herein incorporated by reference into the specification to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated herein by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following Claims.

International Application No: PCT/ /

MICROORGANISMS

Optional Sheet in connection with the microorganism referred to on pages ___, lines _____ of the description ¹

A. IDENTIFICATION OF DEPOSIT ¹

Further deposits are identified on an additional sheet ¹

Name of depositary institution ¹

American Type Culture Collection

Address of depositary institution (including postal code and country) ¹

10801 University Blvd.
Manassas, VA 20110-2209
US

Date of deposit ¹ March 30, 1999 Accession Number ¹ 207178

B. ADDITIONAL INDICATIONS ¹ (leave blank if not applicable). This information is continued on a separate attached sheet

C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE ¹ (if the indications are not all designated States)

D. SEPARATE FURNISHING OF INDICATIONS ¹ (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later ¹ (Specify the general nature of the indications e.g., "Accession Number of Deposit")

E. ☑ This sheet was received with the International application when filed (to be checked by the receiving Office)

(Authorized Officer)

☐ The date of receipt (from the applicant) by the International Bureau ¹

was

(Authorized Officer)

Form PCT/RO/134 (January 1981)

International Application No: PCT/ /

Form PCT/RO/134 (cont.)

American Type Culture Collection

10801 University Blvd.
Manassas, VA 20110-2209
US

Accession No.	Date of Deposit
PTA-249	June 18, 1999
PTA-250	June 18, 1999

What is claimed is:

1. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, 30, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207178, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-249, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-250, or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 300 nucleotides of the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, 30, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207178, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-249, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-250, or a complement thereof;

c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, 29, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207178, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-249, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-250;

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, 29, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207178, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-249, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-250, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, 29, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207178, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-249, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-250; and

e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20,

23, 26, 29, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207178, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-249, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-250, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, 30, or a complement thereof, under stringent conditions.

2. The isolated nucleic acid molecule of Claim 1, which is selected from the group consisting of:

a) a nucleic acid comprising the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, 30, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207178, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-249, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-250, or a complement thereof; and

b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, 29, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207178, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-249, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-250.

3. The nucleic acid molecule of Claim 1 further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of Claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell which contains the nucleic acid molecule of Claim 1.

6. The host cell of Claim 5 which is a mammalian host cell.

7. A non-human mammalian host cell containing the nucleic acid molecule of Claim 1.

8. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207178, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-249, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-250, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NOs: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, or a complement thereof under stringent conditions; and

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, 30, or a complement thereof.

9. The isolated polypeptide of Claim 8 comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29.

10. The polypeptide of Claim 8 further comprising heterologous amino acid sequences.

11. An antibody which selectively binds to a polypeptide of Claim 8.

12. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207178, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-249, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-250;

b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207178, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-249, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-250, wherein the

fragment comprises at least 15 contiguous amino acids of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207178, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-249, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-250; and

c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NOs:2, 5, 8, 11, 14, 17, 20, 23, 26, or 29, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207178, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-249, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-250, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NOs:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 22, 24, 25, 27, 28, 30, or a complement thereof under stringent conditions;

comprising culturing the host cell of Claim 5 under conditions in which the nucleic acid molecule is expressed.

13. A method for detecting the presence of a polypeptide of Claim 8 in a sample, comprising:

a) contacting the sample with a compound which selectively binds to a polypeptide of Claim 8; and

b) determining whether the compound binds to the polypeptide in the sample.

14. The method of Claim 13, wherein the compound which binds to the polypeptide is an antibody.

15. A kit comprising a compound which selectively binds to a polypeptide of Claim 8 and instructions for use.

16. A method for detecting the presence of a nucleic acid molecule of Claim 1 in a sample, comprising the steps of:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

17. The method of Claim 16, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

18. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of Claim 1 and instructions for use.

19. A method for identifying a compound which binds to a polypeptide of Claim 8 comprising the steps of:

a) contacting a polypeptide, or a cell expressing a polypeptide of Claim 8 with a test compound; and

b) determining whether the polypeptide binds to the test compound.

20. The method of Claim 19, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detecting of test compound/polypeptide binding;

b) detection of binding using a competition binding assay;

c) detection of binding using an assay for INTERCEPT 340-, MANGO 003-, MANGO 347-, TANGO 272-, TANGO 295-, TANGO 354-, or TANGO 378-mediated signal transduction.

21. A method for modulating the activity of a polypeptide of Claim 8 comprising contacting a polypeptide or a cell expressing a polypeptide of Claim 8 with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

22. A method for identifying a compound which modulates the activity of a polypeptide of Claim 8, comprising:

a) contacting a polypeptide of Claim 8 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

GTCGACCCACGCGTCCGTTATGTAACTATACATTTTCCCAGAAATTTTAGTATATGATATGATTTTGTTTTCTTTCATC 79

CCTTTTCCCAAGCAGTTTATTATGAAAATTTTCAAACATACAGCAATGTTGAGAAAATTTTACAGTAAATGCCTATACC 158

CATTACCTAAATTTTACCATTAACATTTTACCCTGCTGGCATTATTGTGCTTATCCATCTACGTATCCCTCTCTCCCTT 237

CATTGGTGTATTTCTAAGTAAATTGTAGGCCTCAGTACACTTCCTTCTGAATTCTTCAGCATGCACAACAGTATTATAT 316

TCCATTTTTAAAAGAGCAATTCTTGATAGATTTATATAGTTTTGTAAAATGTTCATATAGAGCTACAAATTTTATCTTT 395

TTGTTTCTTATTGTATGTCTAGGGTCCTGAAGGGGATGCTGGCATTGTTGGGATATCAGGTCCTAAAGGTCCTATTGGA 474

CACAGAGGAAACACTGGTCCCCTTGGCAGAGAAGGTATAATAGGCCCAACAGGTAGAACTGGACCCAGAGGTGAAAAGG 553

GCTTTAGAGGTGAAACTGGTCCTCAAGGACCAAGAGGTCAACCAGGGCCTCCAGGTCCACCTGGAGCACCAGGCCCAAG 632

AAAGCAAATGGATATCAATGCTGCTATTCAAGCCTTGATTGAATCAAATACTGCCCTACAGATGGAGGTAACATATCTG 711

GTTTTATTTATATTGGCACTGTCTCTCAATATACCAATTAAACAGAGAAAATTTTTGGAGGCCAAAATGTGACATTATC 790

TCAAAGATTGTATTTAAAACAGATTGAAAATGTGAAACCATTCTCAAGAACAAAGTAAGTGATTTTGGTATAATTAAAC 869

AGAAATATATGCGTAGGATGTTTTGTAAGGAAAACATTTAAATCAAAAATTTAGTACTGTTATTTGTAAGGAATTTGGT 948

ACTATCCAAGAAAGTAGTTAAATGAGGTTAGCCATGTTTCTTAAAATGAGATATATATATTATCACTACTCATTTATTT 1027

AAACTCTAATGATTCAATGTGTAATTTAAAAAACATAATACAGTAGACATAGCAATTCTTATGTTAGCTTGAAAACTAA 1106

ACTTGCAAATGTGAATTTAACCTCTTTAAAAGATTAAGGTTATTAAAGCATACACATATGCCTATGCTTAAATATAAAC 1185

M E T H S S P A L A 10
TGTTCTTTACATTCTACTCACAACTTACTACACATA ATG GAA ACA CAT TCT TCT CCT GCC TTG GCC 1251

H V G P Q D F F V Y I I L M M T W Q S Y 30
CAT GTT GGT CCT CAG GAT TTT TTT GTT TAT ATA ATT CTT ATG ATG ACT TGG CAG AGC TAC 1311

Q N T E V T L I D H S E E I F K T L N Y 50
CAG AAT ACT GAA GTG ACT TTA ATT GAC CAC AGT GAA GAG ATA TTC AAA ACC CTG AAC TAC 1371

L S N L L H S I K N P L G T R D N P A R 70
CTT AGC AAT TTA TTG CAC AGC ATC AAG AAT CCT CTT GGC ACA CGA GAT AAC CCA GCA CGA 1431

I C K D L L N C E Q K V S D G K Y W I D 90
ATC TGC AAA GAT TTA CTT AAC TGT GAA CAA AAA GTA TCA GAT GGA AAA TAC TGG ATT GAC 1491

P N L G C P S D A I E V F C N F S A G G 110
CCA AAT CTT GGC TGT CCT TCA GAT GCC ATT GAG GTT TTC TGC AAT TTC AGT GCT GGT GGC 1551

Q T C L P P V S V T K L E F G V G K V Q 130
CAG ACA TGC TTA CCT CCT GTT TCT GTA ACA AAG TTG GAG TTT GGA GTT GGG AAA GTC CAG 1611

M N F L H L L S S E A T H I I T I H C L 150
ATG AAC TTC CTT CAT TTA CTG AGT TCG GAA GCC ACC CAT ATC ATC ACC ATT CAC TGT CTA 1671

Figure 1A

```
 N   T   P   R   W   T   S   T   Q   T   S   G   P   G   L   P   I   G   F   K   170
AAC ACC CCA AGG TGG ACA AGC ACA CAA ACA AGT GGC CCA GGA TTG CCT ATT GGT TTC AAG  1731

 G   W   N   G   Q   I   F   K   V   N   T   L   L   E   P   K   V   L   S   D   190
GGA TGG AAT GGC CAG ATT TTT AAA GTA AAC ACT CTA CTT GAA CCT AAA GTG CTT TCA GAT  1791

 D   C   K   I   Q   D   G   S   W   H   K   A   T   F   L   F   H   T   Q   E   210
GAC TGC AAG ATT CAA GAT GGC AGC TGG CAT AAG GCA ACA TTT CTT TTT CAC ACC CAG GAA  1851

 P   N   Q   L   P   V   I   E   V   Q   K   L   P   H   L   K   T   E   R   K   230
CCT AAT CAA CTT CCA GTG ATT GAA GTA CAA AAA CTT CCT CAT CTC AAA ACT GAA CGA AAG  1911

 Y   Y   I   D   S   S   S   V   C   F   L   *                                   242
TAT TAC ATT GAC AGC AGT TCT GTA TGC TTT CTG TAA                                  1947
                                         1944

AGTCTCTGAATTAGTTCCGAATTCAGGCTGTTGGCCAGGTAATTGCTGCAGAGGGAGAAATAAGACAGACAGATACAGT  2026
CATTATGAAATGCATGTAATAAAGCATTGGCTAAATCTTAAAGAATCTCAGGAAGAACAGACTTCCTCCTAAGAAGGAG  2105
AAAAGGCATTTTTAAAGGACTATGATTGATAAAGTATTTAATTCTTTTAAAAATTATATTCATCTCAGCTTTCTTAGAG  2184
AATTCCCTAGAACTAAAAATTTATAAATATGGAATTCTTCAGGGTATCTTATATTTTTGACTGAGTGCGTAGTACCCAT  2263
TAGACAGCTGGAGATGCAGAGCACTATGGAGCAATACTGGCTAATGCTTCCAGATGTGCACTGCTTCTGTCTAAAAATT  2342
ACAAGCCACAGTCTAATATGTCTTATTTTCCAAAACACTAAGCTGTATTCAGGTCCCCGATGGGCATATACATCTTAGC  2421
CGGTGATACACTACCTCTTACGTGTTGCCTCTTTGTGTTGCTTGGTGCTCTTTCGAAAACAAGGTGCTTATGGCTTTCA  2500
TAGACTATTTCCTTTTTCATCTTTGTCATTCTTTAAAAGTGTATGTACTGGTTACATCAAGATATGTTTTGGTTGTTAG  2579
TACTTATTTTAATTTGTTTGGTCACACACTTAATAACACATGAAACTATTTATGTGAAGTCCTTGTTTTATTTTAAAAT  2658
TCTCTTTGTGTATTTGGAATCAAAGCCAGCACATTGTAACCTGTGCTTGTACGCAAAAGAATTAGATTTCTTTGTTTTT  2737
GTTTTATTTTTTAAATTGTTGTAAAAATTATTATAGGCCAGCTACATCTAGTAGTAGGTTTGGGGTACAGATTGGGGGT  2816
TGTGCCATACTGTTTTTAAAGTTCATGATCATCTGGAATGATACTTAGTGTATATATATTTTGTAAAGTTTTAATTCAG  2895
CAAATTTTTTGAAATTGCTGCTGTTTTAAATTATAAAACCTTTATATTTCTGCTTTGTAGAAATTATATGTTTTGTAGT  2974
ATTCATTGATTTTCTTTCACTGTACTTAAATTTAGTGTTAGTACTTTAAAATTTTTAATTTACCAGTCTTTAAAGCAAC  3053
ATCCAGAAAAAAAAAAGTCTTTTCCCATTTAAAATAGGCTCAGCCAGTTCAATGTCGCCTTGTTATCAGAGAAATATTA  3132
GTTCAATACTGAAAGAAAAATATTATACCTCTTGGTATCTAGAAAAGCTTGTTCATCCATTATAAATATATCTTTAGCC  3211
ACAGCAAACCACACTTAACCTATCTATAATAAAAATGTGCTTTAAATAAAAAAAAAAAAAAAAAAGGGCGGCCG       3284
```

Figure 1B

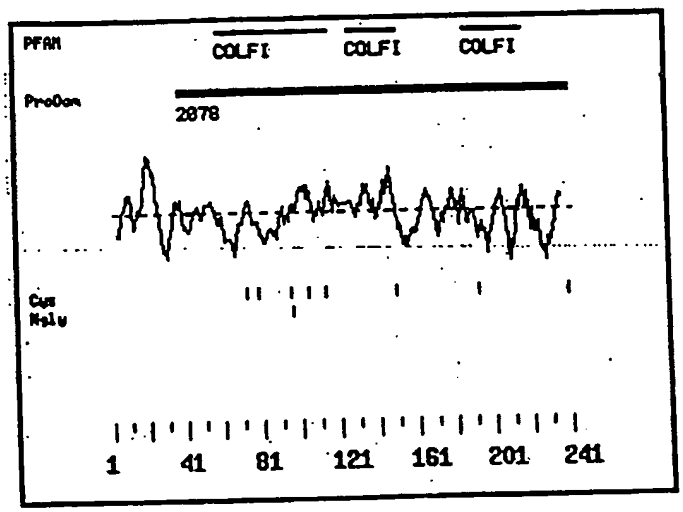
PFAM
COLFI
COLFI
COLFI
ProDom
2878
Cys
Nglu
1
41
81
121
161
201
241

Figure 2

```
    *->lksPeGksrknPARtCkDLfLchpefksGeYWiDPNqGCikDAikVf
       +k+P+G +r+nPAR CkDL  c +  ++G YWiDPN+GC+ DAi+Vf
58     IKNPLG-TRDNPARICKDLLNCEQKVSDGKYWIDPNLGCPSDAIEVF 103

    CnkrfetGvgeTCisp<-*
    Cn  f +G g+TC +p
104 CN--FSAG-GQTCLPP    116


    *->isnvQlTFLRLLSteAsQNiTYhCKN<-*
       +++VQ+ FL LLS+eA   iT hC N
126    VGKVQMNFLHLLSSEATHIITIHCLN    151


    *->tvlGeDGCssrtgewgKTViEyeTkKttRLPIv<-*
       +vl  D C+    g w K+ + + T+ + +LP +
186    KVL-SDDCKIQDGSWHKATFLFHTQEPNQLPVI    217
```

Figure 3

```
                                                         M   T   P   S   P          5
GTCGACCCACGCGTCCGCGCCCGCTGAGCCCCCCGCCGAGGTCCGGACAGGCCGAG ATG ACG CCG AGC CCC       71

 L   L   L   L   L   L   P   P   L   L   L   G   A   F   P   P   A   A   A   A     25
CTG TTG CTG CTC CTG CTG CCG CCG CTG CTG CTG GGG GCC TTC CCG CCG GCC GCC GCC GCC   131

 R   G   P   P   K   M   A   D   K   V   V   P   R   Q   V   A   R   L   G   R     45
CGA GGC CCC CCA AAG ATG GCG GAC AAG GTG GTC CCA CGG CAG GTG GCC CGG CTG GGC CGC   191

 T   V   R   L   Q   C   P   V   E   G   D   P   P   P   L   T   M   W   T   K     65
ACT GTG CGG CTG CAG TGC CCA GTG GAG GGG GAC CCG CCG CCG CTG ACC ATG TGG ACC AAG   251

 D   G   R   T   I   H   S   G   W   S   R   F   R   V   L   P   Q   G   L   K     85
GAT GGC CGC ACC ATC CAC AGC GGC TGG AGC CGC TTC CGC GTG CTG CCG CAG GGG CTG AAG   311

 V   K   Q   V   E   R   E   D   A   G   V   Y   V   C   K   A   T   N   G   F    105
GTG AAG CAG GTG GAG CGG GAG GAT GCC GGC GTG TAC GTG TGC AAG GCC ACC AAC GGC TTC   371

 G   S   L   S   V   N   Y   T   L   V   V   L   D   D   I   S   P   G   K   E    125
GGC AGC CTG AGC GTC AAC TAC ACC CTC GTC GTG CTG GAT GAC ATT AGC CCA GGG AAG GAG   431

 S   L   G   P   D   S   S   S   G   G   Q   E   D   P   A   S   Q   Q   W   A    145
AGC CTG GGG CCC GAC AGC TCC TCT GGG GGT CAA GAG GAC CCC GCC AGC CAG CAG TGG GCA   491

 R   P   R   F   T   Q   P   S   K   M   R   R   R   V   I   A   R   P   V   G    165
CGA CCG CGC TTC ACA CAG CCC TCC AAG ATG AGG CGC CGG GTG ATC GCA CGG CCC GTG GGT   551

 S   S   V   R   L   K   C   V   A   S   G   H   P   R   P   D   I   T   W   M    185
AGC TCC GTG CGG CTC AAG TGC GTG GCC AGC GGG CAC CCT CGG CCC GAC ATC ACG TGG ATG   611

 K   D   D   Q   A   L   T   R   P   E   A   A   E   P   R   K   K   K   W   T    205
AAG GAC GAC CAG GCC TTG ACG CGC CCA GAG GCC GCT GAG CCC AGG AAG AAG AAG TGG ACA   671

 L   S   L   K   N   L   R   P   E   D   S   G   K   Y   T   C   R   V   S   N    225
CTG AGC CTG AAG AAC CTG CGG CCG GAG GAC AGC GGC AAA TAC ACC TGC CGC GTG TCG AAC   731

 R   A   G   A   I   N   A   T   Y   K   V   D   V   I   Q   R   T   R   S   K    245
CGC GCG GGC GCC ATC AAC GCC ACC TAC AAG GTG GAT GTG ATC CAG CGG ACC CGT TCC AAG   791

 P   V   L   T   G   T   H   P   V   N   T   T   V   D   F   G   G   T   T   S    265
CCC GTG CTC ACA GGC ACG CAC CCC GTG AAC ACG ACG GTG GAC TTC GGG GGG ACC ACG TCC   851

 F   Q   C   K   V   R   S   D   V   K   P   V   I   Q   W   L   K   R   V   E    285
TTC CAG TGC AAG GTG CGC AGC GAC GTG AAG CCG GTG ATC CAG TGG CTG AAG CGC GTG GAG   911

 Y   G   A   E   G   R   H   N   S   T   I   D   V   G   G   Q   K   F   V   V    305
TAC GGC GCC GAG GGC CGC CAC AAC TCC ACC ATC GAT GTG GGC GGC CAG AAG TTT GTG GTG   971

 L   P   T   G   D   V   W   S   R   P   D   G   S   Y   L   N   K   L   L   I    325
CTG CCC ACG GGT GAC GTG TGG TCG CGG CCC GAC GGC TCC TAC CTC AAT AAG CTG CTC ATC  1031

 T   R   A   R   Q   D   D   A   G   M   Y   I   C   L   G   A   N   T   M   G    345
ACC CGT GCC CGC CAG GAC GAT GCG GGC ATG TAC ATC TGC CTT GGC GCC AAC ACC ATG GGC  1091
```

Figure 4A

Y S F R S A F L T V L P D P K P P G P P 365
TAC AGC TTC CGC AGC GCC TTC CTC ACC GTG CTG CCA GAC CCA AAA CCG CCA GGG CCA CCT 1151

V A S S S S A T S L P W P V V I G I P A 385
GTG GCC TCC TCG TCC TCG GCC ACT AGC CTG CCG TGG CCC GTG GTC ATC GGC ATC CCA GCC 1211

G A V F I L G T L L L W L C Q A Q K K P 405
GGC GCT GTC TTC ATC CTG GGC ACC CTG CTC CTG TGG CTT TGC CAG GCC CAG AAG AAG CCG 1271

C T P A P A P P L P G H R P P G T A R D 425
TGC ACC CCC GCG CCT GCC CCT CCC CTG CCT GGG CAC CGC CCG CCG GGG ACG GCC CGC GAC 1331

R S G D K D L P S L A A L S A G P G V G 445
CGC AGC GGA GAC AAG GAC CTT CCC TCG TTG GCC GCC CTC AGC GCT GGC CCT GGT GTG GGG 1391

L C E E H G S P A A P Q H L L G P G P V 465
CTG TGT GAG GAG CAT GGG TCT CCG GCA GCC CCC CAG CAC TTA CTG GGC CCA GGC CCA GTT 1451

A G P K L Y P K L Y T D I H T H T H T H 485
GCT GGC CCT AAG TTG TAC CCC AAA CTC TAC ACA GAC ATC CAC ACA CAC ACA CAC ACA CAC 1511

S H T H S H V E G K V H Q H I H Y Q C * 505
TCT CAC ACA CAC TCA CAC GTG GAG GGC AAG GTC CAC CAG CAC ATC CAC TAT CAG TGG TAG 1571

1565

ACGGCACCGTATCTGCAGTGGGCACGGGGGGGCCGGCCAGACAGGCAGACTGGGAGGATGGAGGACGGAGCTGCAGACG 1650

AAGGCAGGGGACCCATGGCGAGGAGGAATGGCCAGCACCCCAGGCAGTCTGTGTGTGAGGCATAGCCCCTGGACACACA 1729

CACACAGACACACACACTGCCTGGATGCATGTATGCACACACATGCGCGCACACGTGCTCCCTGAAGGCACACGTACGC 1808

ACACACGCACATGCACAGATATGCCGCCTGGGCACACAGATAAGCTGCCCAAATGCACGCACACGCACAGAGACATGCC 1887

AGAACATACAAGGACATGCTGCCTGAACATACACACGCACACCCATGCGCAGATGTGCTGCCTGGACACACACACACAC 1966

ACGGATATGCTGTCTGGACGCACACACGTGCAGATATGGTATCCGGACACACACGTGCACAGATATGCTGCCTGGACAC 2045

ACAGATAATGCTGCCTTGACACACACATGCACGGATATTGCCTGGACACACACACACACACGTGTGCACAGATATGCTG 2124

TCTGGACACGCACACACATGCAGATATGCTGCCTGGACACACACTTCCAGACACACGTGCACAGGCGCAGATATGCTGC 2203

CTGGACACACGCAGATATGCTGTCTAGTCACACACACACGCAGACATGCTGTCCGGACACACACACGCATGCACAGATA 2282

TGCTGTCCGGACACACACACGCACGCAGATATGCTGCCTGGACACACACACAGATAATGCTGCCTCAACACTCACACAC 2361

GTGCAGATATTGCCTGGACACACACATGTGCACAGATATGCTGTCTGGACATGCACACACGTGCAGATATGCTGTCCGG 2440

ATACACACGCACGCACACATGCAGATATGCTGCCTGGGCACACACTTCCGGACACACATGCACACACAGGTGCAGATAT 2519

GCTGCCTGGACACACGCAGACTGACGTGCTTTTGGGAGGGTGTGCCGTGAAGCCTGCAGTACGTGTGCCGTGAGGCTCA 2598

TAGTTGATGAGGGACTTTCCCTGCTCCACCGTCACTCCCCCAACTCTGCCCGCCTCTGTCCCCGCCTCAGTCCCCGCCT 2677

CCATCCCCGCCTCTGTCCCCTGGCCTTGGCGGCTATTTTTGCCACCTGCCTTGGGTGCCCAGGAGTCCCCTACTGCTGT 2756

GGGCTGGGGTTGGGGGCACAGCAGCCCCAAGCCTGAGAGGCTGGAGCCCATGGCTAGTGGCTCATCCCCACTGCATTCT 2835

CCCCCTGACACAGAGAAGGGGCCTTGGTATTTATATTTAAGAAATGAAGATAATATTAATAATGATGGAAGGAAGACTG 2914

Figure 4B

GGTTGCAGGGACTGTGGTCTCTCCTGGGGCCCGGACCCGCCTGGTCTTTCAGCCATGCTGATGACCACACCCCGTCCA 2993

GGCCAGACACCACCCCCCACCCCACTGTCGTGGTGGCCCCAGATCTCTGTAATTTTATGTAGAGTTTGAGCTGAAGCCC 3072

CGTATATTTAATTTATTTTGTTAAACATGAAAGTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3151

AAAAAAAAGGGCGGCCGC 3169

Figure 4C

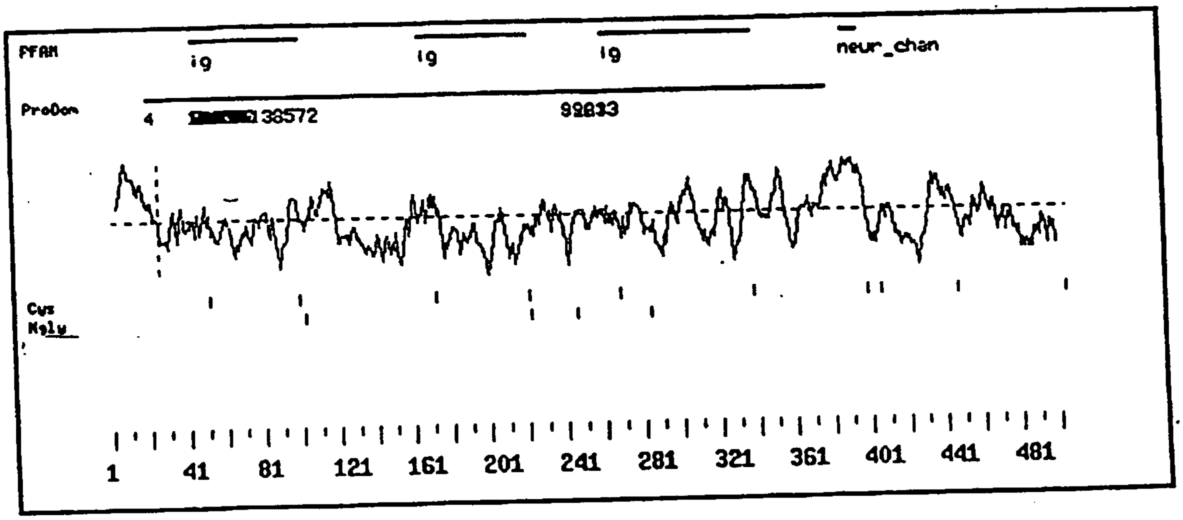
PFAM
ig
ig
ig
neur_chan
ProDom
4
38572
99833
Cys
Ngly
1
41
81
121
161
201
241
281
321
361
401
441
481

Figure 5

```
               *->GesvtLtCsvsgfgpp.p.vtWlrngk.............lslti.s
                  G +v+L+C v   g+p+p   W+++g++ +++ ++ +   +   l ++
M003     44       GRTVRLQCPVE--GDPpPlTMWTKDGRtihsgwsrfrvlpQGLKVkQ 88

               vtpeDsgGtYtCvv<-*
               v++eD+ G+Y C +
M003     89 VEREDA-GVYVCKA    101

               *->GesvtLtCsvsgfgpp.p.vtWlrngk..............lslti.
                  G+sv+L C +s  g p+p++tW ++++  ++++  ++++++ +l ++
M003    165       GSSVRLKCVAS--GHPrPdITWMKDDQaltrpeaaeprkkkWTLSLk 209

               svtpeDsgGtYtCvv<-*
               +++peDs G YtC+v
M003    210 NLRPEDS-GKYTCRV    223

               *->GesvtLtCsvsgfgpp.p.vtWlrngk....................
                  G++ +++C v+  ++ +p ++Wl+  + + +++++++ + ++++
M003    261       GGTTSFQCKVR--SDVkPvIQWLKRVEygaegrhnstidvggqkfvv 305

               ...............lslti.svtpeDsgGtYtCvv<-*
                ++++  ++++++    l+i++++++D+ G Y C
M003    306 lptgdvwsrpdgsylNKLLItRARQDDA-GMYICLG    340
```

Figure 6

```
              *->vfvlGTlgif<-*
                 vf+lGTl ++
M003    388      VFILGTLLLW    397
```

Figure 7

R V R P T G D V W S R P D G S Y L N K 19
CA CGC GTC CGG CCC ACG GGT GAT GTG TGG TCA CGG CCT GAT GGC TCC TAC CTC AAC AAG 59

L L I S R A R Q D D A G M Y I C L G A N 39
CTG CTC ATC TCT CGG GCC CGC CAG GAT GAT GCT GGC ATG TAC ATC TGC CTA GGT GCA AAT 119

T M G Y S F R S A F L T V L P D P K P P 59
ACC ATG GGC TAC AGT TTC CGT AGC GCC TTC CTC ACT GTA TTA CCA GAC CCC AAA CCT CCA 179

G P P M A S S S S S T S L P W P V V I G 79
GGG CCT CCT ATG GCT TCT TCA TCG TCA TCC ACA AGC CTG CCA TGG CCT GTG GTG ATC GGC 239

I P A G A V F I L G T V L L W L C Q T K 99
ATC CCA GCT GGT GCT GTC TTC ATC CTA GGC ACT GTG CTG CTC TGG CTT TGC CAG ACC AAG 299

K K P C A P A S T L P V P G H R P P G T 119
AAG AAG CCA TGT GCC CCA GCA TCT ACA CTT CCT GTG CCT GGG CAT CGT CCC CCA GGG ACA 359

S R E R S G D K D L P S L A V G I C E E 139
TCC CGA GAA CGC AGT GGT GAC AAG GAC CTG CCC TCA TTG GCT GTG GGC ATA TGT GAG GAG 419

H G S A M A P Q H I L A S G S T A G P K 159
CAT GGA TCC GCC ATG GCC CCC CAG CAC ATC CTG GCC TCT GGC TCA ACT GCT GGC CCC AAG 479

L Y P K L Y T D V H T H T H T H T C T H 179
CTG TAC CCC AAG CTA TAC ACA GAT GTG CAC ACA CAC ACA CAT ACA CAC ACC TGC ACT CAC 539

T L S C W R A R F I N T S M S T I S A K 199
ACG CTC TCA TGT TGG AGG GCA AGG TTC ATC AAC ACC AGC ATG TCC ACT ATC AGT GCT AAA 599

Y S E S P S T V S * 209
TAC AGC GAA TCT CCA AGC ACT GTG TCC TGA 629

GGTAGGCATTTGGGGGCCAAGGCAACAGGTTGGGAGAATTGAGAACAATGGAGGAAGAGTATCTTAGGGTGCCTTATGG 708

TGGACACTCACAAACTTGGCCATATAGATGTATGTACTACCAGATGAACAGCCAGCCAGATTCACACACGCACATGTTT 787

AAACGTGTAAACGTGTGCACAACTGCACACACAACCTGAGAAACCTTCAGGAGGATTTGTGGTGTGACTTTGCAGTGAC 866

ATGTAGCGATGGCTAGTTGAAGGAATCTCCCTCATGTCTTAGTGGTCATGGCCACTTCCCCACCCCTGCCCATCTGTGT 945

TCCTGCCTGGCCTTGGTGGTGCTTCCGTGTGCCCTGGGTTTTCCAGGAACCCTATCAACCTGACTGGGGTGAGCAGTGC 1024

AGCCATGCNTGGAGGTTTGAGCCACCCTCCCCTTGCTAGAGAGAAGGGCN 1074

Figure 8

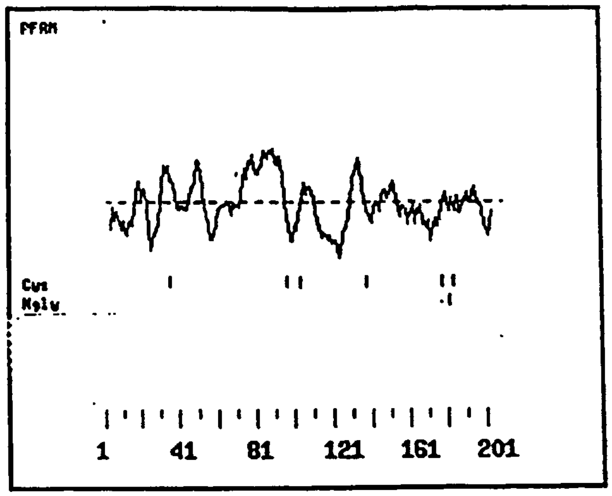

Figure 9

M P G P R V W G K Y L W 12
GTCGACCCACGCGTCCGCCCACGCGTCCGG ATG CCT GGA CCC AGA GTG TGG GGG AAA TAT CTC TGG 66

R S P H S K G C P G A M W W L L L W G V 32
AGA AGC CCT CAC TCC AAA GGC TGT CCA GGC GCA ATG TGG TGG CTG CTT CTC TGG GGA GTC 126

L Q A C P T R G S V L L A Q E L P Q Q L 52
CTC CAG GCT TGC CCA ACC CGG GGC TCC GTC CTC TTG GCC CAA GAG CTA CCC CAG CAG CTG 186

T S P G Y P E P Y G K G Q E S S T D I K 72
ACA TCC CCC GGG TAC CCA GAG CCG TAT GGC AAA GGC CAA GAG AGC AGC ACG GAC ATC AAG 246

A P E G F A V R L V F Q D F D L E P S Q 92
GCT CCA GAG GGC TTT GCT GTG AGG CTC GTC TTC CAG GAC TTC GAC CTG GAG CCG TCC CAG 306

D C A G D S V T V S W G W G G S R Q D C 112
GAC TGT GCA GGG GAC TCT GTC ACA GTG AGC TGG GGA TGG GGG GGG TCC CGC CAG GAC TGT 366

G Q G D S R G C G K W R C P E S P I W R 132
GGC CAG GGA GAT TCC CGG GGT TGT GGG AAG TGG CGG TGC CCT GAA TCC CCC ATC TGG AGG 426

R D E F S M * 139
AGG GAT GAA TTT TCC ATG TAG 447

GGGCAGTCGGGCTTGGCTTACCGGGGAGCAGTGGTGGACCCCAGGACACAGCCTCCCACCAGCGCCTCCGGGGCTGCCA 526

TCTGGGCCCCACAGAGCAAAGAGGGCAGCAAGCAGGCCCTGCGTTTGGAAGGCTTATGAATGGACACACAAATCTTGCA 605

AATCTATGGAGCCAGGGGCAGGGACGCACATATTGGTTGTTAAAAATATGTCATCATGTATTTGTTGAGTGCCTGCTCT 684

ATCAGGTGAGGAAGCTGGACACAAATAATAACAAAAGATTAAGTCACCGTTCACACTTACCTTGGAAGAGCTATTACAA 763

AACTTCTAACGCCAAAGCCTTATTCAGAATAAGGACATTTTAAAAACAGTACTTGATGGAGTGATGCAAGCTTGCAGTC 842

CCAGCAGTATAGTCAGGAGACTGAGGCTGGAGGATCAGAGGGCTGGAGCCCAGGGTTCAAGGCCAGCCTAAGCAACATA 921

GCAAGACCCCATCTCAAAAATAAGTAAATAATAAATAAAAATAAAAAGAGCACATTATCTTTTGATTTAAATTTTATTT 1000

ATATCAAAATGACATAAATTTTTGAACTTTATTTTTTAATTTTAAAATTTTTAATTATTATGGATACATAATAGTTGTA 1079

AGACTTTTTGTTTTTTAATTAAAGTTTTCTAAGGCTGGGCGCAGTAGCTCATGTCTGTAGTCCCAGCACTTTGGGAGGC 1158

TGAGGCGAAAGAAGCACTTGAGCCCAGGAATTTGAGACCAGCCTGGGCAACATAGCAAGACCCCATCTCTACAAAAAAA 1237

TTTAAAAATTAGCCAAGTGTGGTGGCACGCACCTGTGGTCCCAGCTACAAGGGACGCTGAAGTGAGAGGATCACTTGAG 1316

CCTGGAAGGTAGAGGCTGCAGTGAGCTCTGATCATGACACCGTACTCCAGCCTGGGTGACAGAGTGAGACCCTGTCTCC 1395

AAAAAAAAAAAAAAAAAAGGGCGGCCGC 1423

Figure 10

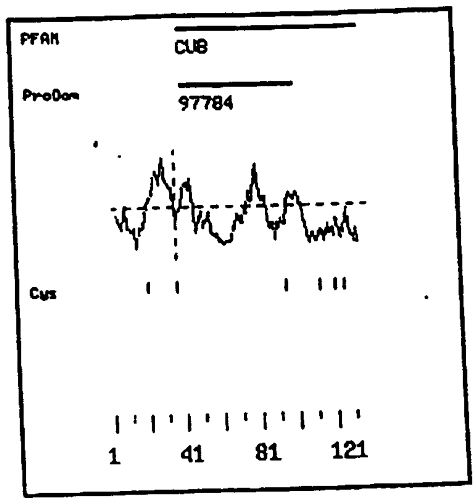

Figure 11

```
    *->CGgtldltessGsisSPnYPnrsdYppnkeCvWrIrappgyrvVeLt
        G +l+ +e +  ++SP+YP+  +Y   +e    I ap+g+  V L
 40    -GSVLLAQELPQQLTSPGYPE--PYGKGQESSTDIKAPEGFA-VRLV 82

       FqdFdlEdhdgapCryDyvEirDGdpss.pllG....rfCG....sgkPe
       FqdFdlE +++   C+ D+v +  G ++s++  G++++r CG+ +   ++P
 83    FQDFDLEPSQD--CAGDSVTVSWGWGGSrQDCGqgdsRGCGkwrcPESP- 129

       dirStsnrmlikFvsDasvskrGFkAty<-*
                   +  +D+          +
130    ------------IWRRDE-----------F    136
```

Figure 12

```
GTCGACCCACGCGTCCGCTCGAAGCGGGGACCCTCGCCCCGTCCTCGGCTGTCCAGTCCTCCTCCTCGCAGACCCCGGC  79
GGTTCCTACCCCAGGCCGCAGGGGAGACGGTGCCCCAAGGCAGGCTTCATATCCTGAACGCTGGGATCCCCCAGGACAT  158
                                                                     M   S      2
TCCCTGGCCCCCAGGCCCCAGGTCCCAGGCCCCAGGGCTGAGCTGTGGGCAGGCCCCACCTGGCCTCTGCA ATG TCA  235
 P   P   L   C   P   L   L   L   L   A   V   G   L   R   L   A   G   T   L   N      22
CCG CCT CTG TGT CCC CTC CTT CTC CTG GCT GTG GGC CTG CGG CTG GCT GGA ACT CTC AAC    295
 P   S   D   P   N   T   C   S   F   W   E   S   F   T   T   T   T   K   E   S      42
CCC AGT GAT CCC AAT ACC TGC AGC TTC TGG GAA AGC TTC ACT ACC ACC ACC AAG GAG TCC    355
 H   S   R   P   F   S   L   L   P   S   E   P   C   E   R   P   W   E   G   P      62
CAC TCC CGC CCC TTC AGC CTG CTC CCC TCA GAG CCC TGC GAG CGG CCC TGG GAG GGC CCC    415
 H   T   C   P   S   P   Q   T   Q   R   K   L   L   A   S   R   D   S   F   C      82
CAT ACT TGC CCC AGC CCA CAA ACT CAG AGG AAA CTC CTG GCT TCT AGG GAT TCA TTC TGC    475
 M   V   C   V   G   A   G   V   Q   W   R   D   R   S   A   L   Q   P   Q   T     102
ATG GTC TGT GTC GGG GCT GGA GTG CAG TGG CGA GAT CGT AGT GCA CTG CAA CCT CAA ACA    535
 G   N   A   L   S   M   R   P   Q   P   R   V   L   S   G   A   P   S   L   A     122
GGG AAT GCG CTT TCT ATG CGC CCT CAG CCC AGA GTG TTG AGT GGT GCC CCT TCC CTG GCC    595
 S   P   G   H   T   V   V   V   K   T   D   H   R   Q   R   L   Q   C   C   H     142
TCC CCT GGC CAC ACT GTG GTG GTG AAG ACG GAC CAC CGC CAG CGC CTG CAG TGC TGC CAT    655
 G   F   Y   E   S   R   G   F   C   V   P   L   C   A   Q   E   C   V   H   G     162
GGC TTC TAT GAG AGC AGG GGG TTC TGT GTC CCG CTC TGT GCC CAG GAG TGT GTC CAT GGC    715
 R   C   V   A   P   N   Q   C   Q   C   V   P   G   W   R   G   D   D   C   S     182
CGT TGT GTG GCA CCC AAT CAG TGC CAA TGT GTG CCA GGC TGG CGG GGC GAC GAC TGT TCC    775
 S   A   P   N   C   L   Q   P   C   T   P   G   Y   Y   G   P   A   C   Q   F     202
AGT GCC CCG AAC TGC CTT CAG CCC TGT ACC CCT GGC TAC TAT GGC CCT GCC TGC CAG TTC    835
 R   C   Q   C   H   G   A   P   C   D   P   Q   T   G   A   C   F   C   P   A     222
CGC TGC CAG TGC CAT GGG GCA CCC TGC GAT CCC CAG ACT GGA GCC TGC TTC TGC CCC GCA    895
 E   R   T   G   P   S   C   D   V   S   C   S   Q   G   T   S   G   F   F   C     242
GAG AGA ACT GGG CCC AGC TGT GAC GTG TCC TGT TCC CAG GGC ACT TCT GGC TTC TTC TGC    955
 P   S   T   H   P   C   Q   N   G   G   V   F   Q   T   P   Q   G   S   C   S     262
CCC AGC ACC CAT CCT TGC CAA AAT GGA GGT GTC TTC CAA ACC CCA CAG GGC TCC TGC AGC   1015
 C   P   P   G   W   M   G   T   I   C   S   L   P   C   P   E   G   F   H   G     282
TGC CCC CCT GGC TGG ATG GGC ACC ATC TGC TCC CTG CCC TGC CCA GAG GGC TTT CAC GGA   1075
 P   N   C   S   Q   E   C   R   C   H   N   G   G   L   C   D   R   F   T   G     302
CCC AAC TGC TCC CAG GAA TGT CGC TGC CAC AAC GGC GGC CTC TGT GAC CGA TTC ACT GGG   1135
 Q   C   R   C   A   P   G   Y   T   G   D   R   C   R   E   E   C   P   V   G     322
CAG TGC CGC TGC GCT CCG GGT TAC ACT GGG GAT CGG TGC CGG GAG GAG TGC CCG GTG GGC   1195
```

Figure 13A

R F G Q D C A E T C D C A P D A R C F P 342
CGC TTT GGG CAG GAC TGT GCT GAG ACG TGC GAC TGC GCC CCG GAC GCC CGT TGC TTC CCG 1255

A N G A C L C E H G F T G D R C T D R L 362
GCC AAC GGC GCA TGT CTG TGC GAA CAC GGC TTC ACT GGG GAC CGC TGC ACG GAT CGC CTC 1315

C P D G F Y G L S C Q A P C T C D R E H 382
TGC CCC GAC GGC TTC TAC GGT CTC AGC TGC CAG GCC CCC TGC ACC TGC GAC CGG GAG CAC 1375

S L S C H P M N G E C S C L P G W A G L 402
AGC CTC AGC TGC CAC CCG ATG AAC GGG GAG TGC TCC TGC CTG CCG GGC TGG GCG GGC CTC 1435

H C N E S C P Q D T H G P G C Q E H C L 422
CAC TGC AAC GAG AGC TGC CCG CAG GAC ACG CAT GGG CCA GGG TGC CAG GAG CAC TGT CTC 1495

C L H G G V C Q A T S G L C Q C A P G Y 442
TGC CTG CAC GGT GGC GTC TGC CAG GCT ACC AGC GGC CTC TGT CAG TGC GCG CCG GGT TAC 1555

T G P H C A S L C P P D T Y G V N C S A 462
ACG GGC CCT CAC TGT GCT AGT CTT TGT CCT CCT GAC ACC TAC GGT GTC AAC TGT TCT GCA 1615

R C S C E N A I A C S P I D G E C V C K 482
CGC TGC TCA TGT GAA AAT GCC ATC GCC TGC TCA CCC ATC GAC GGC GAG TGC GTC TGC AAG 1675

E G W Q R G N C S V P C P P G T W G F S 502
GAA GGT TGG CAG CGT GGT AAC TGC TCT GTG CCC TGC CCA CCC GGA ACC TGG GGC TTC AGT 1735

C N A S C Q C A H E A V C S P Q T G A C 522
TGC AAT GCC AGC TGC CAG TGT GCC CAT GAG GCA GTC TGC AGC CCC CAA ACT GGA GCC TGT 1795

T C T P G W H G A H C Q L P C P K G Q F 542
ACC TGC ACC CCT GGG TGG CAT GGG GCC CAC TGC CAG CTG CCC TGT CCG AAG GGG CAG TTT 1855

G E G C A S R C D C D H S D G C D P V H 562
GGA GAA GGT TGT GCC AGT CGC TGT GAC TGT GAC CAC TCT GAT GGC TGT GAC CCT GTT CAT 1915

G R C Q C Q A G W M G A R C H L S C P E 582
GGA CGC TGT CAG TGC CAG GCT GGC TGG ATG GGT GCC CGC TGC CAC CTG TCC TGC CCT GAG 1975

G L W G V N C S N T C T C K N G G T C L 602
GGC TTA TGG GGA GTC AAC TGT AGC AAC ACC TGC ACC TGC AAG AAT GGG GGC ACC TGT CTC 2035

P E N G N C V C A P G F R G P S C Q R S 622
CCT GAG AAT GGC AAC TGC GTG TGT GCA CCC GGA TTC CGG GGC CCC TCC TGC CAG AGA TCC 2095

C Q P G R Y G K R C V P C K C A N H S F 642
TGT CAG CCT GGC CGC TAT GGC AAA CGC TGT GTG CCC TGC AAG TGC GCT AAC CAC TCC TTC 2155

C H P S N G T C Y C L A G W T G P D C S 662
TGC CAC CCC TCG AAC GGG ACC TGC TAC TGC CTG GCT GGC TGG ACA GGC CCC GAC TGC TCC 2215

Q P C P P G H W G E N C A Q T C Q C H H 682
CAG CCA TGC CCT CCA GGA CAC TGG GGA GAA AAC TGT GCC CAG ACC TGC CAA TGT CAC CAT 2275

G G T C H P Q D G S C I C P L G W T G H 702
GGT GGG ACC TGC CAT CCC CAG GAT GGG AGC TGT ATC TGC CCC CTA GGC TGG ACT GGA CAC 2335

Figure 13B

H C L E G C P L G T F G A N C S Q P C Q 722
CAC TGC TTA GAA GGC TGC CCT CTG GGG ACA TTT GGT GCT AAC TGC TCC CAG CCA TGC CAG 2395

C G P G E K C H P E T G A C V C P P G H 742
TGT GGT CCT GGA GAA AAG TGC CAC CCA GAG ACT GGG GCC TGT GTA TGT CCC CCA GGG CAC 2455

S G A P C R I G I Q E P F T V M P T T P 762
AGT GGT GCA CCT TGC AGG ATT GGA ATC CAG GAG CCC TTT ACT GTG ATG CCG ACC ACT CCA 2515

V A Y N S L G A V I G I A V L G S L V V 782
GTA GCG TAT AAC TCG CTG GGT GCA GTG ATT GGC ATT GCA GTG CTG GGG TCC CTT GTG GTA 2575

A L V A L F I G Y R H W Q K G K E H H H 802
GCC CTG GTG GCA CTG TTC ATT GGC TAT CGG CAC TGG CAA AAA GGC AAG GAG CAC CAC CAC 2635

L A V A Y S S G R L D G S E Y V M P D V 822
CTG GCT GTG GCT TAC AGC AGC GGG CGC CTG GAC GGC TCC GAG TAT GTC ATG CCA GAT GTC 2695

P P S Y S H Y Y S N P S Y H T L S Q C S 842
CCT CCG AGC TAC AGT CAC TAC TAC TCC AAC CCC AGC TAC CAC ACC CTG TCG CAG TGC TCC 2755

P N P P P P N K V P G P L F A S L Q N P 862
CCA AAC CCC CCA CCC CCT AAC AAG GTT CCA GGC CCG CTC TTT GCC AGC CTG CAG AAC CCT 2815

E R P G G A Q G H D N H T T L P A D W K 882
GAG CGG CCA GGT GGG GCC CAA GGG CAT GAT AAC CAC ACC ACC CTG CCT GCT GAC TGG AAG 2875

H R R E P P P G P L D R G S S R L D R S 902
CAC CGC CGG GAG CCC CCT CCA GGG CCT CTG GAC AGG GGG AGC AGC CGC CTG GAC CGA AGC 2935

Y S Y S Y S N G P G P F Y D K G L I S E 922
TAC AGC TAT AGC TAC AGC AAT GGC CCA GGC CCA TTC TAC GAT AAA GGG CTC ATC TCT GAA 2995

E E L G A S V A S L S S E N P Y A T I R 942
GAG GAG CTC GGG GCC AGT GTG GCT TCC CTG AGC AGT GAG AAC CCA TAT GCC ACC ATC CGG 3055

D L P S L P G G P R E S S Y M E M K G P 962
GAC CTG CCC AGC TTG CCA GGG GGC CCC CGG GAG AGC AGC TAC ATG GAG ATG AAA GGC CCT 3115

P S G S A P R Q P P Q F W D S Q R R R Q 982
CCC TCA GGA TCT GCC CCC AGG CAG CCT CCT CAG TTT TGG GAC AGC CAG AGG CGG CGG CAA 3175

P Q P Q R D S G T Y E Q P S P L I H D R 1002
CCC CAG CCA CAG AGA GAC AGT GGC ACC TAC GAG CAG CCC AGC CCC CTG ATC CAT GAC CGA 3235

D S V G S Q P P L P P G L P P G H Y D S 1022
GAC TCT GTG GGC TCC CAG CCC CCT CTG CCT CCG GGC CTA CCC CCC GGC CAC TAT GAC TCA 3295

P K N S H I P G H Y D L P P V R H P P S 1042
CCC AAG AAC AGC CAC ATC CCT GGA CAT TAT GAC TTG CCT CCA ~~GTA CGG CAT CCC~~ CCA TCA 3355

P P L R R Q D R * 1051
CCT CCA CTT CGA CGC CAG GAC CGT TGA 3382

GGAGCCAGGATGGTATGGCAGAGGCCAGCACACCTGGCTGTTGCTGCTCAAGGCTGGGGACAGAGCCTAGTGTACCCCT 3461

Figure 13C

GCCAGGAGCAGGGAGTGGACCGGCAGGCTGTGAACATGAACAACGCTTAACAGAGCAAGTGATGGGAGCCTTGTTCCTG 3540
GGTTCTACCATGGGAGACGCTGATCAGCAGGATGCCTGGCTCCCTTTCCCAACCCACTGCTCCCAAGGCCTCCAGGGCC 3619
CTGTGTACATAAACTGGTGGGTTGGAAGTTGCTGGGTAACTCTGATTTCAGACATGCGTGTGGGGTACCTTTTCTGTGC 3698
ATGCTCAGCCTGGGCTCTGTGCGTGTGTGTGTTTCTGTGATTTTAGAAGGGTACCAGGCAGGTTCTGTCCTAGGGCACT 3777
TACCATTTAGTAGGGAGATGGAACCAACCCAATTAACTCTAGCAATAGCCTCCTAACTGGCCTCCTCCATTGATTCAGT 3856
GAACCTTCCAATGCATGGCTCATAATTTCAAAATACAGGCTGGTTAGTTACTCCCTACCTGAAAGCCTTCATAGGTGCC 3935
TCTTTGCTCTTCTGCCAGTATCAAAACTTTTGAAGGCCTTAAAGGCCCTGCTTTGCCTGGCCCATCTGTCTCTCCAGCC 4014
TCACCTTGAACTGTGTTCCTGTCACTGCACGCCAGTCACACCGGCCTCTAGGTCCTCCTGTAGGCCACTCTTCTTTCTG 4093
GCACAGGGACCTGCACACCTGGAGTGCCCTTCCTCCCCCACTCGCCTGTTCACCCCTGCTTTTCCTTTACACCTCCTCC 4172
TCAGGGAAGTGCCCACCCTCCGTACATCTTTCACAGCCCTGATTGCAGCTGTGTTCACTCACCAGGTACCTGCAGAAGG 4251
CCTACAGGGTGCCAGGCACTTCTTTAATGGGTTCTTTCTTTATGTGATTATTTGATTAATCTCTGCCTCCCCCACTAGA 4330
CTGTAAGCTCCCTGAAGGCAAGAATCCTGTGCTTATGCTCAATATTAGCTCTCCCTTGGCACAGAGTAGGCACTCAACA 4409
AATGCTCCCCAAAAGGCTGAGTGGCTGACTGAATTAAGTACCAGTGACATGCAGTAACTGCTAAGATAGATGAGCCATC 4488
TGTATGCTCTGACAGTTACAGACTGAATAAGTTGGAGACTTCCCTAAAGGGTGGCATTTCCCCAGGGTAACAACGCAGA 4567
GCTCAGGTGTGGGAAGGTGCCAGGGGCAGGGGTGCAGAGGGGCTGAGGCTGAGGGGGGTGCAGAGGCTGGAGAAAGGAT 4646
AACAGGAGAGAGTATACAGGCATGCCTTGATTTATTGCACTTCACAGGTAGCAGAATTTTTAAAGAAATTGAAGGTTTT 4725
GGGACATATATGTGACAGCAATAGGTTAAGAAAAGCAAAGCAGAGAAATTGAAGATTTGTGTCAACACTGCTTTAAGCA 4804
AATCTGTTGGCACCATTTTTCCAATAGCATGTGCCCATTTTGGGTCTCTACATTGCATTTTGGTAATTGCTTGCAATAT 4883
TTCAAGCATTTTCATTGTTATTATATGTGTTATAGTGATCTGTGATCAGTGATCTTTGATATATTATTGTAATTGTTTC 4962
GGGGCGCCATGAACCGCACCCATATAACACGGTAAACTTAATCAGCAAAAAAAAAAAAAAAAAAAGGGCGGCCG 5036

Figure 13D

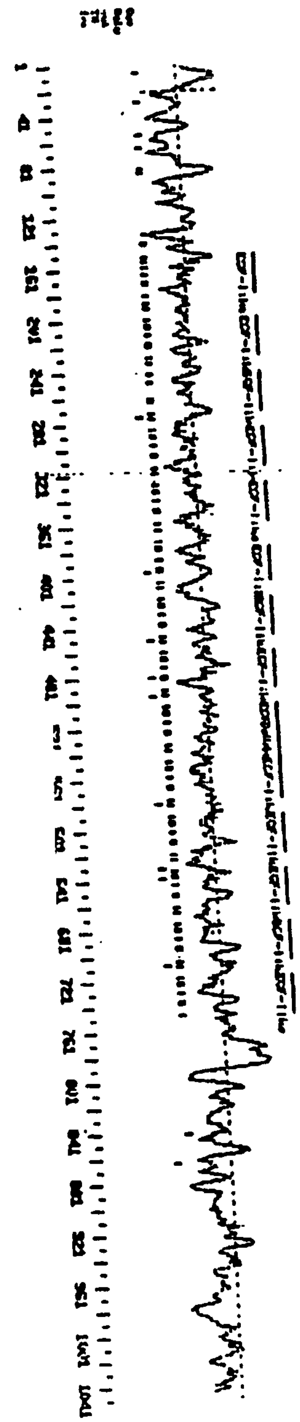

Figure 14

```
    *->Capnn..pCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC
       C p++ + C + G+Cv                +C+C pG      + G++C
151    CVPLCaqECVH-GRCVAPN----------QCQCVPG-----WRGDDC 181

    <-*


    *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
       C+   + C++ + C + g            C+Cp         tG+ C
200    CQFRCQCHG-APCDPQTG---------ACFCPAE-----RTGPSC   229

    *


    *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
       C+++ pC+ngG+ +    g          +C CppG       + G  C
242    CPSTHPCQNGGVFQTPQG---------SCSCPPG-----WMGTIC   272

    *


    *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
       C++++ C+ngG C     g          +C+C+pG       ytG+rC
285    CSQECRCHNGGLCDRFTG---------QCRCAPG-----YTGDRC   315

    *


    *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
       Ca+++ C   +++C +  g            C C +G       +tG+rC
328    CAETCDCAPDARCFPANG---------ACLCEHG-----FTGDRC   358

    *


    *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
       C+ +      +    C++  g           +C C pG       ++G +C
378    CDRE----HSLSCHPMNG---------ECSCLPG-----WAGLHC   404

    *
```

Figure 15A

```
         *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
            C++++ C++gG+C+ t g          C+C+pG     ytG++C
417         CQEHCLCLHGGVCQATSG---------LCQCAPG-----YTGPHC     447

         *

       -    -

         *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
            C+   + C n   C +  g         +C+C++G      ++   +C
460         CSARCSCENAIACSPIDG---------ECVCKEG-----WQRGNC     490

         *

       -    -

         *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
            C+   + C  +  ++C +  g         C+C+pG      ++G  +C
503         CNASCQCAHEAVCSPQTG---------ACTCTPG-----WHGAHC     533

         *

       -    -

         *->WstdkhiggrtslGfnleyrirvtCdenYYGegCnkFCrPrdDafgH
             +t + + + +      + +     C +  +GegC+  C+         H
518         -QTGACTCTPG----WHGAHCQLPCPKGQFGEGCASRCDCD-----H 554

         yt.Cd.enGnklCleGWkGeyC<-*
         + +Cd+ +G+ +C +GW+G  C
555      SDgCDpVHGRCQCQAGWMGARC     576

         *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
            Ca+ + C++    C +++g         +C+C+ G       + G rC
546         CASRCDCDHSDGCDPVHG---------RCQCQAG-----WMGARC     576

         *

       -    -

         *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
            C+ ++ C+ngGtC++   g          C+C+pG      + G+ C
589         CSNTCTCKNGGTCLPENG---------NCVCAPG-----FRGPSC     619

         *

       -    -

         *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
            C p   C n+ +C+++ g          tC C  G       +tG++C
632       - CVPC-KCANHSFCHPSNG---------TCYCLAG-----WTGPDC     661

         *

       -    -
```

Figure 15B

```
     *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
        Ca+++ C++gGtC++  g         +C+Cp G     +tG++C
674     CAQTCQCHHGGTCHPQDG---------SCICPLG-----WTGHHC     704

     *

     -    -


     *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
        C++++ C  g +C++  g          C+CppG      +G  C
717     CSQPCQCGPGEKCHPETG---------ACVCPPG-----HSGAPC     747

     *

     -    -
```

Figure 15C

```
  S   T   H   A   S   G   D   P   V   H   G   Q   C   R   C   Q   A   G   W      19
G TCG ACC CAC GCG TCC GGT GAC CCT GTT CAT GGA CAG TGC CGA TGT CAG GCT GGT TGG    58

M   G   T   R   C   H   L   P   C   P   E   G   F   W   G   A   N   C   S   N    39
ATG GGC ACA CGC TGC CAC CTG CCT TGC CCG GAG GGC TTT TGG GGA GCC AAC TGC AGT AAC 118

T   C   T   C   K   N   G   G   T   C   V   S   E   N   G   N   C   V   C   A    59
ACC TGT ACC TGC AAG AAT GGT GGT ACC TGT GTG TCT GAG AAT GGC AAC TGC GTG TGC GCA 178

P   G   F   R   G   P   S   C   Q   R   P   C   P   P   G   R   Y   G   K   R    79
CCA GGG TTC CGA GGC CCC TCC TGC CAG AGG CCC TGC CCG CCT GGT CGC TAT GGC AAA CGC 238

C   V   Q   C   K   C   N   N   N   H   S   S   C   H   P   S   D   G   T   C    99
TGT GTG CAA TGC AAG TGT AAC AAC AAC CAT TCT TCC TGC CAC CCA TCG GAC GGG ACC TGC 298

S   C   L   A   G   W   T   G   P   D   C   S   E   A   C   P   P   G   H   W   119
TCC TGC CTG GCG GGC TGG ACA GGC CCT GAC TGC TCC GAG GCA TGT CCC CCA GGC CAC TGG 358

G   L   K   C   S   Q   L   C   Q   C   H   H   G   G   T   C   H   P   Q   D   139
GGA CTC AAA TGC TCC CAA CTC TGC CAG TGT CAT CAT GGT GGG ACC TGC CAC CCC CAG GAT 418

G   S   C   I   C   T   P   G   W   T   G   P   N   C   L   E   G   C   P   P   159
GGG AGC TGT ATC TGC ACG CCA GGC TGG ACT GGA CCC AAC TGC TTG GAA GGC TGC CCA CCA 478

R   M   F   G   V   N   C   S   Q   L   C   Q   C   D   L   G   E   M   C   H   179
AGA ATG TTT GGT GTC AAC TGC TCC CAG CTA TGT CAG TGT GAT CTC GGA GAG ATG TGC CAC 538

P   E   T   G   A   C   V   C   P   P   G   H   S   G   A   D   C   K   M   G   199
CCA GAG ACT GGG GCT TGT GTC TGT CCC CCA GGA CAC AGT GGT GCA GAC TGC AAA ATG GGA 598

S   Q   E   S   F   T   I   M   P   T   S   P   V   T   H   N   S   L   G   A   219
AGC CAG GAG TCC TTC ACC ATA ATG CCC ACC TCT CCC GTG ACC CAT AAC TCA CTG GGT GCA 658

V   I   G   I   A   V   L   G   T   L   V   V   A   L   I   A   L   F   I   G   239
GTG ATT GGC ATT GCA GTA CTG GGA ACC CTC GTG GTG GCC CTG ATA GCA CTG TTC ATT GGC 718

Y   R   Q   W   Q   K   G   K   E   H   E   H   L   A   V   A   Y   S   T   G   259
TAC CGC CAG TGG CAA AAG GGC AAG GAA CAT GAG CAC TTG GCA GTG GCT TAC AGC ACT GGG 778

R   L   D   G   S   D   Y   V   M   P   D   V   S   P   S   Y   S   H   Y   Y   279
CGG CTG GAT GGC TCT GAT TAC GTC ATG CCA GAT GTC TCT CCG AGC TAT AGT CAC TAC TAC 838

S   N   P   S   Y   H   T   L   S   Q   C   S   P   N   P   P   P   P   N   K   299
TCC AAC CCC AGC TAC CAC ACA CTG TCT CAG TGT TCT CCT AAC CCC CCG CCC CCT AAC AAG 898

V   P   G   S   Q   L   F   V   S   S   Q   A   P   E   R   P   S   R   A   H   319
GTC CCA GGC AGT CAG CTC TTT GTC AGC TCT CAG GCC CCT GAG CGG CCA AGC AGA GCC CAC 958

G   R   E   N   H   T   T   L   P   A   D   W   K   H   R   R   E   P   H   D   339
GGG CGT GAG AAC CAT ACC ACA CTG CCC GCT GAC TGG AAG CAC CGC CGG GAG CCC CAT GAC 1018

R   G   A   S   H   L   D   R   S   Y   S   C   S   Y   S   H   R   N   G   P   359
AGA GGC GCC AGC CAC CTG GAC CGA AGC TAT AGC TGT AGC TAT AGC CAC AGG AAT GGC CCA 1078
```

Figure 16A

G P F C H K G P I S E E G L G A S V M S 379
GGA CCA TTC TGT CAT AAA GGT CCC ATC TCT GAA GAG GGA CTA GGG GCA AGC GTT ATG TCC 1138

L S S E N P Y A T I R D L P S L P G E P 399
CTG AGC AGT GAG AAC CCC TAT GCT ACC ATC CGA GAC CTG CCC AGC CTG CCT GGG GAA CCC 1198

R E S G Y V E M K G P P S V S P P R Q S 419
CGA GAA AGT GGC TAT GTG GAG ATG AAA GGA CCT CCA TCA GTG TCC CCT CCC AGG CAG TCT 1258

L H L R D R Q Q R Q L Q P Q R D S G T Y 439
CTT CAT CTC CGG GAC AGG CAG CAG CGG CAA CTG CAG CCA CAG AGG GAC AGC GGC ACC TAT 1318

E Q P S P L S H N E E S L G S T P P L P 459
GAG CAG CCC AGC CCC TTG AGC CAT AAT GAA GAG TCT TTG GGC TCC ACG CCC CCG CTT CCT 1378

P G L P P G H Y D S P K N S H I P G H Y 479
CCA GGC CTG CCT CCT GGT CAC TAC GAC TCC CCC AAG AAC AGC CAT ATC CCT GGA CAC TAT 1438

D L P P V R H P P S P P S R R Q D R * 498
GAC TTG CCT CCA GTA CGG CAT CCT CCA TCC CCT CCA TCC CGG CGC CAG GAC CGC TGA 1495

AGAGCCGGCATGGTATGGAGCGTGCCTATGTACCTTGCCAGGAGCAGGGACTGGACCAGCAGGCCACGAACAGAAACA 1574

CTTGGTGAAGTGAACAGAGACGGACTGTGGCCCTGTGCTTCCACCGAGGGAGACACTAGTTGACAAAGTGTCTAACCCT 1653

CTTTTCCAACCCACTGCTCAAGTCCCTGTGGACATAAGCTGGTGGGCAGAATGTTGTTGTACAAGTGTGATTTTAGATC 1732

GATTTTTTTTTAAAGTATGTGTTGGGTACCTTTTCTGTGTGTATGCTCAGGCAGGCTGTGTGTGTCTCTAGTTGGCTTT 1811

AGAGGGAGTCAGGTATAGGTTCTGCCTTCTGCACTTTCCATCTTATCTAGTAGTCAGCTTCCAAGCTTAACTAGTTAGA 1890

GCTCCACCAGCAGCAGGCCCTAACTACCTGCCTGCCCTTCACCCAGTAATCCTCCATGTCTTTGCTCAGAGGATTGCTC 1969

CCCGACTCTGGTGTTGTCCTCCTGGTACGCCTTGACGGTCCTGCAGTCTCCCTTTCCCGTCTTGCTTCATTCTTTCCCA 2048

GAATGAAGGCTGTCTGCCACCCTACTTCCCAGCCCAGGAATTGGCACATCTAAGTTCAGCCTTCCTAAGTTACCCGTTG 2127

AGTCCTGCTTGCCCTTCACATATTCCACAGAACACCCACCCCACATCTGCTTCATAGCTACTCTCTTCTCCACGTACCC 2206

ACAGAAGGCAGAAGTGGTACCAGGCAAGAAGATGGGATTGTTGCATTTTGTTTTGTTTTTGAGACTCTGTCTCACTATG 2285

TAGTCCTGGCTGGCCTGGAACTCAAGAGCTCTGCCTGCCTCTGCCTCTTGAGTGCTGGGTTTAACGGCTCAGGGTCACA 2364

TGCACAGCTCAAGCTGCACTCCGATGTGCTTTCCCCTGTTGCTAGATTAGCGTCTGCCTCCCCCTAGTGGAGAGGCTGA 2443

TCGCCAGCTCTCTGATGCAGGACTCTGGTGTTTAGGCTCACTCACTATTGGTTTCCTTGGCACAGGGTAGTCACTCAAT 2522

AAATGTTCCTCTAAAAGCTGAAAAAAAAAAAAAAAAAGGGCGGCCGC 2569

Figure 16B

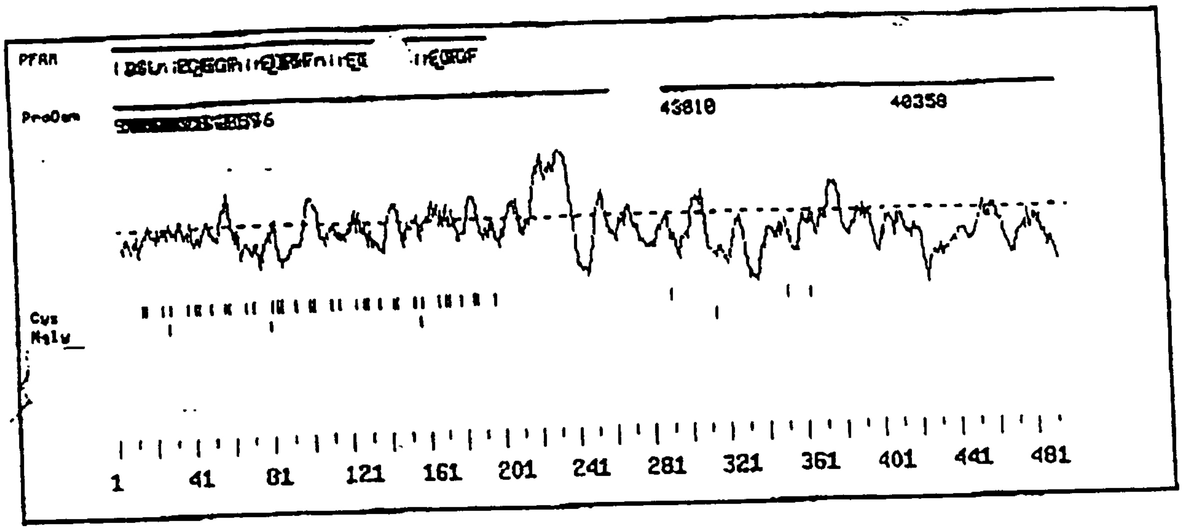
PFAM
ProDom
43810
40358
Cys
Hslw
1
41
81
121
161
201
241
281
321
361
401
441
481

Figure 17

```
GTCGACCCACGCGTCCGGCTCCCAGCCCACCCCCAAACAGACACAGCGTAGCCCGGGCCAGCTCTTAAGGAGTTCAGGA  79
GTGAGAAGAGGCCCTCAGAGATCTGACAGCCTAGGAGTGCGTGGACACCACCTCAGCCCACTGAGCAGGAGTCACAGCA  158
                                                  M   A   P   A   R                5
CGAAGACCAAGCGCAAAGCGACCCCTGCCCTCCATCCTGACTGCTCCTCCTAAGAGAG ATG GCA CCG GCC AGA   231

 A   G   F   C   P   L   L   L   L   L   L   L   G   L   W   V   A   E   I   P    25
GCA GGA TTC TGC CCC CTT CTG CTG CTT CTG CTG CTG GGG CTG TGG GTG GCA GAG ATC CCA   291

 V   S   A   K   P   K   G   M   T   S   S   Q   W   F   K   I   Q   H   M   Q    45
GTC AGT GCC AAG CCC AAG GGC ATG ACC TCA TCA CAG TGG TTT AAA ATT CAG CAC ATG CAG   351

 P   S   P   Q   A   C   N   S   A   M   K   N   I   N   K   H   T   K   R   C    65
CCC AGC CCT CAA GCA TGC AAC TCA GCC ATG AAA AAC ATT AAC AAG CAC ACA AAA CGG TGC   411

 K   D   L   N   T   F   L   H   E   P   F   S   S   V   A   A   T   C   Q   T    85
AAA GAC CTC AAC ACC TTC CTG CAC GAG CCT TTC TCC AGT GTG GCC GCC ACC TGC CAG ACC   471

 P   K   I   A   C   K   N   G   D   K   N   C   H   Q   S   H   G   P   V   S   105
CCC AAA ATA GCC TGC AAG AAT GGC GAT AAA AAC TGC CAC CAG AGC CAC GGG CCC GTG TCC   531

 L   T   M   C   K   L   T   S   G   K   Y   P   N   C   R   Y   K   E   K   R   125
CTG ACC ATG TGT AAG CTC ACC TCA GGG AAG TAT CCG AAC TGC AGG TAC AAA GAG AAG CGA   591

 Q   N   K   S   Y   V   V   A   C   K   P   P   Q   K   K   D   S   Q   Q   F   145
CAG AAC AAG TCT TAC GTA GTG GCC TGT AAG CCT CCC CAG AAA AAG GAC TCT CAG CAA TTC   651

 H   L   V   P   V   H   L   D   R   V   L   *                                   157
CAC CTG GTT CCT GTA CAC TTG GAC AGA GTC CTT TAG                                  687

GTTTCCAGACTGGCTTGCTCTTTGGCTGACCTTCAATTCCCTCTCCAGGACTCCGCACCACTCCCCTACACCCAGAGCA  766
TTCTCTTCCCCTCATCTCTTGGGGCTGTTCCTGGTTCAGCCTCTGCTGGGAGGCTGAAGCTGACACTCTGGTGAGCTGA  845
GCTCTAGAGGGATGGCTTTTCATCTTTTTGTTGCTGTTTTCCCAGATGCTTATCCCCAAGAAACAGCAAGCTCAGGTCT  924
GTGGGTTCCCTGGTCTATGCCATTGCACATGTCTCCCCTGCCCCCTGGCATTAGGGCAGCATGACAAGGAGAGGAAATA  1003
AATGGAAAGGGGGCATATGGGATTTGTGGACACAGCTGTTTCTGTTCCTGAACTAGAAGTCTTCCCCAGCTCTGACGTG  1082
GCAGTGAGGTGACCTGAAGGAAAGAAAAATATAAATAAATACCACTTCATATTTGTATAGAATCCTCTAATCCCTTGTG  1161
ACATAGACTTGACAGGGATTGTATGCCTTCTTTATGGATGAGGAAATTAAGGTTTTAGAAAGCTTAATGAATTAAAGAG  1240
CTTGTCTAATTAGTTAGTAGCAGAACCTGGACTTGAACCTAGGTCTCCTTGCTCTAAATACAGTGTACCTTCTACTCTA  1319
CCAGTTGCGCAAGAAAGAAGTCACTGTTACAGAGGCAAGCGGTGAACTAGGTAAGAGTTCACTCATGAAGAAACGAGTG  1398
CTCTGAAGAGCCAGTTACCCTGTGTTGGCTGCAATAAAGGTCATTACCTCTCTAGCCAAAAAAAAAAAAAAAAAAAAAA  1477
AAAAAAAAAAAAAAAAAAAA                                                             1497
```

Figure 18

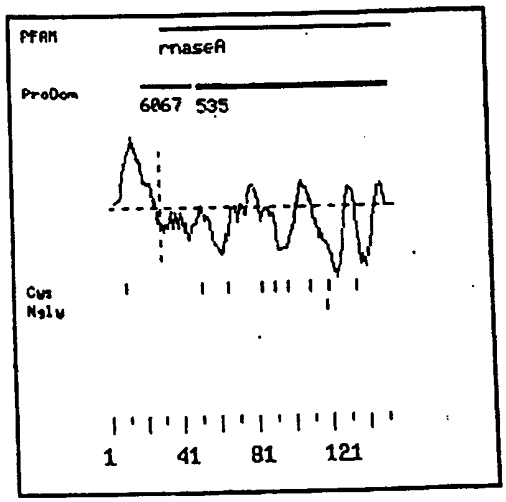

Figure 19

```
       *->qesrAqkFlrQHiDspktsssnpnYCNqMMdkrRnmtqgrCKpvNTF
             + ++ q+F++QH+    ++s +     CN +M k++n     rCK+ NTF
 32          GMTSSQWFKIQHM---QPSPQA---CNSAM-KNINKHTKRCKDLNTF 71

          vHesladVkaVCsqkNvtCkNGqkNCyqSkssfqiTdCrltggsqkyPnC
          +He++++V a C ++ + CkNG kNC+qS+ ++++T C+lt+g   yPnC
 72       LHEPFSSVAATCQTPKIACKNGDKNCHQSHGPVSLTMCKLTSGK--YPNC 119

          rYrtsastkhIiVACEgrd.rddPyynPyvPVHFDasv<-*
          rY+ + ++k ++VAC +++++d+ ++    vPVH+D++
120       RYKEKRQNKSYVVACKPPQkKDSQQFH-LVPVHLDRVL    156
```

Figure 20

```
                                                                 M   P   L   L      4
GTCGACCCACGGCGTCCGGCCAGGCTCCACTGAGGGGAACGGGGACCTGTCTGAAGAGAAG ATG CCC CTG CTG     73

 T   L   Y   L   L   L   F   W   L   S   G   Y   S   I   A   T   Q   I   T   G      24
ACA CTC TAC CTG CTC CTC TTC TGG CTC TCA GGC TAC TCC ATT GCC ACT CAA ATC ACC GGT   133

 P   T   T   V   N   G   L   E   R   G   S   L   T   V   Q   C   V   Y   R   S      44
CCA ACA ACA GTG AAT GGC TTG GAG CGG GGC TCC TTG ACC GTG CAG TGT GTT TAC AGA TCA   193

 G   W   E   T   Y   L   K   W   W   C   R   G   A   I   W   R   D   C   K   I      64
GGC TGG GAG ACC TAC TTG AAG TGG TGG TGT CGA GGA GCT ATT TGG CGT GAC TGC AAG ATC   253

 L   V   K   T   S   G   S   E   Q   E   V   K   R   D   R   V   S   I   K   D      84
CTT GTT AAA ACC AGT GGG TCA GAG CAG GAG GTG AAG AGG GAC CGG GTG TCC ATC AAG GAC   313

 N   Q   K   N   R   T   F   T   V   T   M   E   D   L   M   K   T   D   A   D     104
AAT CAG AAA AAC CGC ACG TTC ACT GTG ACC ATG GAG GAT CTC ATG AAA ACT GAT GCT GAC   373

 T   Y   W   C   G   I   E   K   T   G   N   D   L   G   V   T   V   Q   V   T     124
ACT TAC TGG TGT GGA ATT GAG AAA ACT GGA AAT GAC CTT GGG GTC ACA GTT CAA GTG ACC   433

 I   D   P   A   S   T   P   A   P   T   T   P   T   S   T   T   F   T   A   P     144
ATT GAC CCA GCG TCG ACT CCT GCC CCC ACC ACG CCT ACT TCC ACT ACG TTT ACA GCA CCA   493

 V   T   Q   E   E   T   S   S   S   P   T   L   T   G   H   H   L   D   N   R     164
GTC ACC CAA GAA GAA ACT AGC AGC TCC CCA ACT CTG ACC GGC CAC CAC TTG GAC AAC AGG   553

 H   K   L   L   K   L   S   V   L   L   P   L   I   F   T   I   L   L   L   L     184
CAC AAG CTC CTG AAG CTC AGT GTC CTC CTG CCC CTC ATC TTC ACC ATA TTG CTG CTG CTT   613

 L   V   A   A   S   L   L   A   W   R   M   M   K   Y   Q   Q   K   A   A   G     204
TTG GTG GCC GCC TCA CTC TTG GCT TGG AGG ATG ATG AAG TAC CAG CAG AAA GCA GCC GGG   673

 M   S   P   E   Q   V   L   Q   P   L   E   G   D   L   C   Y   A   D   L   T     224
ATG TCC CCA GAG CAG GTA CTG CAG CCC CTG GAG GGC GAC CTC TGC TAT GCA GAC CTG ACC   733

 L   Q   L   A   G   T   S   P   R   K   A   T   T   K   L   S   S   A   Q   V     244
CTG CAG CTG GCC GGA ACC TCC CCG CGA AAG GCT ACC ACG AAG CTT TCC TCT GCC CAG GTT   793

 D   Q   V   E   V   E   Y   V   T   M   A   S   L   P   K   E   D   I   S   Y     264
GAC CAG GTG GAA GTG GAA TAT GTC ACC ATG GCT TCC TTG CCG AAG GAG GAC ATT TCC TAT   853

 A   S   L   T   L   G   A   E   D   Q   E   P   T   Y   C   N   M   G   H   L     284
GCA TCT CTG ACC TTG GGT GCT GAG GAT CAG GAA CCG ACC TAC TGC AAC ATG GGC CAC CTC   913

 S   S   H   L   P   G   R   G   P   E   E   P   T   E   Y   S   T   I   S   R     304
AGT AGC CAC CTC CCC GGC AGG GGC CCT GAG GAG CCC ACG GAA TAC AGC ACC ATC AGC AGG   973

 P   *                                                                             306
CCT TAG                                                                            979

CCTGCACTCCAGGCTCCTTCTTGGACCCCAGGCTGTGAGCACACTCCTGCCTCATCGACCGTCTGCCCCCTGCTCCCCT   1058

CATCAGGACCAACCCGGGGACTGGTGCCTCTGCCTGATCAGCCAGCATTGCCCCTAGCTCTGGGTTGGGCTTGGGGCCA   1137
```

Figure 21A

AGTCTCAGGGGGCTTCTAGGAGTTGGGGTTTTCTAAACGTCCCCTCCTCTCCTACATAGTTGAGGAGGGGGCTAGGGAT 1216

ATGCTCTGGGGCTTTCATGGGAATGATGAAGATGATAATGAGAAAAATGTTATCATTATTATCATGAAGTACCATTATC 1295

ATAATACAATGAACCTTTATTTATTGCCTACCACATGTTATGGGCTGAATAATGGCCCCCAAAGATATCTGTGTCCTAA 1374

TCCTCAGAACTTGTGACTGTTACCTTCTGTGGCAGAAAGGGACAGTGCAGATGTATGTAAGTTAAGGACTTTGAGATAG 1453

AGAGGTTATTCTTGCTGATTCAGGTGGGCCCAAAATATCACCACAAGGGTCCTCATAAGAAAGAGGCCAGAAGGTCAAA 1532

GAGGTAGAGACAAAGTGATGATGGAAGTGGACGTGGGTGTGACGTGAGCAGGGGCCATGAATGCCGCAGCCTTCAGATG 1611

CCAGAAAGGGAAAGGAATGGATTCCCCTGCCTGGAGCCTCCAAAAGAAACCAGCCCTGCCCACGCCTTGACTTGAGCCC 1690

ATTGAAACTGATCTTGAGCTCCTGGCCTCCAGAATTGCAGGAGAATAAATTTGTGTTGTTTTAAAAAAAAAAAAAAAA 1769

AAAAAGGGCGGCCGCTAGA 1788

Figure 21B

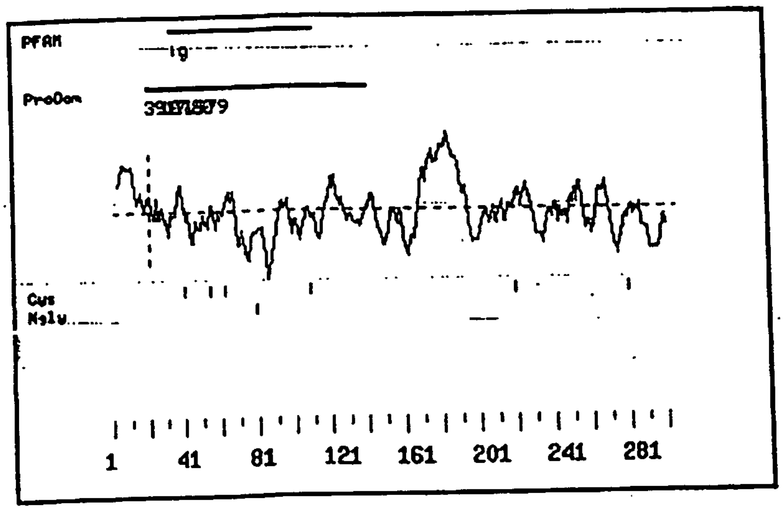

Figure 22

```
   *->GesvtLtCsvsgfgppgvsvtWyf.....kngk.lgpsllgysysrl
      ++s+t +C ++  + + +++ W+ ++   ++ k l    ++ s   +
33    RGSLTVQCVYR--SGWETYLKWWCrgaiwRDCKiLVK--TSGSEQEV 75

   esgekanlsegrfsis......sltLtissvekeDsGtYtCvv<-*
    ++        r+si ++++++++t+t+ ++ k D+ tY+C
76 KRD--------RVSIKdnqknrTFTVTMEDLMKTDADTYWCGI    110
```

Figure 23

```
                                           M   D   H   C   G   A   L   F   L      9
CACGCGTCCGGCCAGTTCTTGGAGGAGACTCTGCACAGGGC ATG GAT CAC TGT GGT GCC CTT TTC CTG     68

 C   L   C   L   L   T   L   Q   N   A   T   T   E   T   W   E   E   L   L   S      29
TGC CTG TGC CTT CTG ACT TTG CAG AAT GCA ACA ACA GAG ACA TGG GAA GAA CTC CTG AGC    128

 Y   M   E   N   M   Q   V   S   R   G   R   S   S   V   F   S   S   R   Q   L      49
TAC ATG GAG AAT ATG CAG GTG TCC AGG GGC CGG AGC TCA GTT TTT TCC TCT CGT CAA CTC    188

 H   Q   L   E   Q   M   L   L   N   T   S   F   P   G   Y   N   L   T   L   Q      69
CAC CAG CTG GAG CAG ATG CTA CTG AAC ACC AGC TTC CCA GGC TAC AAC CTG ACC TTG CAG    248

 T   P   T   I   Q   S   L   A   F   K   L   S   C   D   F   S   G   L   S   L      89
ACA CCC ACC ATC CAG TCT CTG GCC TTC AAG CTG AGC TGT GAC TTC TCT GGC CTC TCG CTG    308

 T   S   A   T   L   K   R   V   P   Q   A   G   G   Q   H   A   R   G   Q   H     109
ACC AGT GCC ACT CTG AAG CGG GTG CCC CAG GCA GGA GGT CAG CAT GCC CGG GGT CAG CAC    368

 A   M   Q   F   P   A   E   L   T   R   D   A   C   K   T   R   P   R   E   L     129
GCC ATG CAG TTC CCC GCC GAG CTG ACC CGG GAC GCC TGC AAG ACC CGC CCC AGG GAG CTG    428

 R   L   I   C   I   Y   F   S   N   T   H   F   F   K   D   E   N   N   S   S     149
CGG CTC ATC TGT ATC TAC TTC TCC AAC ACC CAC TTT TTC AAG GAT GAA AAC AAC TCA TCT    488

 L   L   N   N   Y   V   L   G   A   Q   L   S   H   G   H   V   N   N   L   R     169
CTG CTG AAT AAC TAC GTC CTG GGG GCC CAG CTG AGT CAT GGG CAC GTG AAC AAC CTC AGG    548

 D   P   V   N   I   S   F   W   H   N   Q   S   L   E   G   Y   T   L   T   C     189
GAT CCT GTG AAC ATC AGC TTC TGG CAC AAC CAA AGC CTG GAA GGC TAC ACC CTG ACC TGT    608

 V   F   W   K   E   G   A   R   K   Q   P   W   G   G   W   S   P   E   G   C     209
GTC TTC TGG AAG GAG GGA GCC AGG AAA CAG CCC TGG GGG GGC TGG AGC CCT GAG GGC TGT    668

 R   T   E   Q   P   S   H   S   Q   V   L   C   R   C   N   H   L   T   Y   F     229
CGT ACA GAG CAG CCC TCC CAC TCT CAG GTG CTC TGC CGC TGC AAC CAC CTC ACC TAC TTT    728

 A   V   L   M   Q   L   S   P   A   L   V   P   A   E   L   L   A   P   L   T     249
GCT GTT CTC ATG CAA CTC TCC CCA GCC CTG GTC CCT GCA GAG TTG CTG GCA CCT CTT ACG    788

 Y   I   S   L   V   G   C   S   I   S   I   V   A   S   L   I   T   V   L   L     269
TAC ATC TCC CTC GTG GGC TGC AGC ATC TCC ATC GTG GCC TCG CTG ATC ACA GTC CTG CTG    848

 H   F   H   F   R   K   Q   S   D   S   L   T   R   I   H   M   N   L   H   A     289
CAC TTC CAT TTC AGG AAG CAG AGT GAC TCC TTA ACA CGC ATC CAC ATG AAC CTG CAT GCC    908

 S   V   L   L   L   N   I   A   F   L   L   S   P   A   F   A   M   S   P   V     309
TCC GTG CTG CTC CTG AAC ATC GCC TTC CTG CTG AGC CCC GCA TTC GCA ATG TCT CCT GTG    968

 P   G   S   A   C   T   A   L   A   A   A   L   H   Y   A   L   L   S   C   L     329
CCC GGG TCA GCA TGC ACG GCT CTG GCC GCT GCC CTG CAC TAC GCG CTG CTC AGC TGC CTC   1028

 T   W   M   A   I   E   G   F   N   L   Y   L   L   L   G   R   V   Y   N   I     349
ACC TGG ATG GCC ATC GAG GGC TTC AAC CTC TAC CTC CTC CTC GGG CGT GTC TAC AAC ATC   1088
```

Figure 24A

```
 Y   I   R   R   Y   V   F   K   L   G   V   L   G   W   G   A   P   A   L   L   369
TAC ATC CGC AGA TAT GTG TTC AAG CTT GGT GTG CTA GGC TGG GGG GCC CCA GCC CTC CTG 1148

 V   L   L   S   L   S   V   K   S   S   V   Y   G   P   C   T   I   P   V   F   389
GTG CTG CTT TCC CTC TCT GTC AAG AGC TCG GTA TAC GGA CCC TGC ACA ATC CCC GTC TTC 1208

 D   S   W   E   N   G   T   G   F   Q   N   M   S   I   C   W   V   R   S   P   409
GAC AGC TGG GAG AAT GGC ACA GGC TTC CAG AAC ATG TCC ATA TGC TGG GTG CGG AGC CCC 1268

 V   V   H   S   V   L   V   M   G   Y   G   G   L   T   S   L   F   N   L   V   429
GTG GTG CAC AGT GTC CTG GTC ATG GGC TAC GGC GGC CTC ACG TCC CTC TTC AAC CTG GTG 1328

 V   L   A   W   A   L   W   T   L   R   R   L   R   E   R   A   D   A   P   S   449
GTG CTG GCC TGG GCG CTG TGG ACC CTG CGC AGG CTG CGG GAG CGG GCG GAT GCA CCA AGT 1388

 V   R   A   C   H   D   T   V   T   V   L   G   L   T   V   L   L   G   T   T   469
GTC AGG GCC TGC CAT GAC ACT GTC ACT GTG CTG GGC CTC ACC GTG CTG CTG GGA ACC ACC 1448

 W   A   L   A   F   F   S   F   G   V   F   L   L   P   Q   L   F   L   F   T   489
TGG GCC TTG GCC TTC TTT TCT TTT GGC GTC TTC CTG CTG CCC CAG CTG TTC CTC TTC ACC 1508

 I   L   N   S   L   Y   G   F   F   L   F   L   W   F   C   S   Q   R   C   R   509
ATC TTA AAC TCG CTC TAC GGT TTC TTC CTT TTC CTG TGG TTC TGC TCC CAG CGG TGC CGC 1568

 S   E   A   E   A   K   A   Q   I   E   A   F   S   S   S   Q   T   T   Q   *   529
TCA GAA GCA GAG GCC AAG GCA CAG ATA GAG GCC TTC AGC TCC TCC CAA ACA ACA CAG TAG 1628

TCCGGGCCTCCTGGCCTGGAATCCTCAGCCTCTCTGGCCGCCAGTAGCCTGAGGCTACGGCTCCTGCTAGAGAGGGTGG 1707
CAGGCCTGCTGCTGGACCCCAGAGGCCACTGTGACCGCCAAGGGGCCTTTTCCACTTCCACGGCCTCTCCAGGCACTGA 1786
GGGGAAGGCATTGCTCTACCTCTCCCTGACATTTTGCTCCGGGGCAGATCCAACCTTACCTGGGGCAGCAAACTTTGTC 1865
CTGGTACCTGGGCCCAGCTCGCCAGGGATGTGGCAGAGCACCAGCCTGGGCATCAGGAAGCCAAGTTTCAAGGACTGT 1944
CTTTGAGTCTGTCTGTATGACCTTGGGCCTGCCACTTCTCACAGACCCTAGGTATCCACAGCTGTGACATGGGGGCAAG 2023
CGGCTTTGTTTCAGCCTAACCCAGGAGCTTAGTAAAAATTGCATAAGACCAGGGGGAAGAGTGTCAGCGTGGGGTGGGA 2102
ATTCCCGCGGCCTCCACCTGCTTGCTAGGGGCAGGATCTCATTCAGGCTGCCCTGGAAGCACCTGCTTGGCCCTGCCAC 2181
CTTCCTCCAGGGGAGGGCCAGATGGCATCCTGGCTTGGGGCGGGTGGACCTACCCAGGCTCTGAGACTTTACTGGCCT 2260
ATGCCTGAGGCCTCTTTTCCTTTAACTCCCTAAATTATGATGACTCCAAGTCCAAGCCCACCCTTCCCAAAGATTGGGA 2339
GGTTCCGCCGTTCCCAGAGGCTCCTCCTGCGGTGCTCCCAAGACTTCCATAGACCATCTGGACCAGTAGCCCATCCCGC 2418
AGTTTTCTTGGGGGCAGAGGAAAACGCTTCTTTCTCCTCCAGCTGAATCAGCTGGATCCCAGTGTCCTGGCTGTTTGGT 2497
GATTGGGCAAGATTGAATTTGCCCAGGTAGGCGTGAGAGTGTGGGTTTTAAATTCGAAGCTCAGGCCATAGTTTCAGAG 2576
AATCACCCTTACCCCAGACCTTCATGAGACAGTGCTCATGAAGCCAGTGCGTTTCCCAGAACGAACACTAGGCGGCACC 2655
GTTGGTCCACACTCAGAGGCCCTTGGCGCCAAGACTGCATCTAGAATCGCTCAAACACCTGTTTGCAGACCCCATGCAC 2734
CAGCTGGAGGGGCCGTAACTGCAGGACTGCGCCTACTGAGTGACCCATTTCCTCCAGGAGGAAAGGCAAGACACGCTTA 2813
```

Figure 24B

CACGGCCATTTGTCTCTTTTCCCAATGCGGCGGTGCACTTTCGCTCTTGGGGGCTGCACCCCAGACATAGCTGGCACCA 2892

GAGCAGGGTGCTCAGGTGGTGGGTGCTCAGGGCCCTGCCCCAGGCCACTGGGCCGTTTTGATGACCTCGAAGGTCACAG 2971

GCAGAAAATAGGAGCAGGATTTCCCCTGGGGAAAAGTTCTCCTGGGACATCTTCTGCTCTTCTGTACATTTCTAGATGC 3050

AAATAACTCCTTCACCAGGCAGTGAGTGGCGTAGGCTCTGGAGCCAGGCTGCCTGGGCTCCAATGCCAGCTCTGCCACT 3129

TGCTAGCTGTGAGACTGTGGACAAACCACTCAGCCTCTGTGTGCCTCAGTTTTCCTATTTGTAAAATAGAGGCCATAGT 3208

GGTACCTATTTTGAAGACTAAGTAAAAGAATTCAAATAAAGAGACTTGGC 3258

Figure 24C

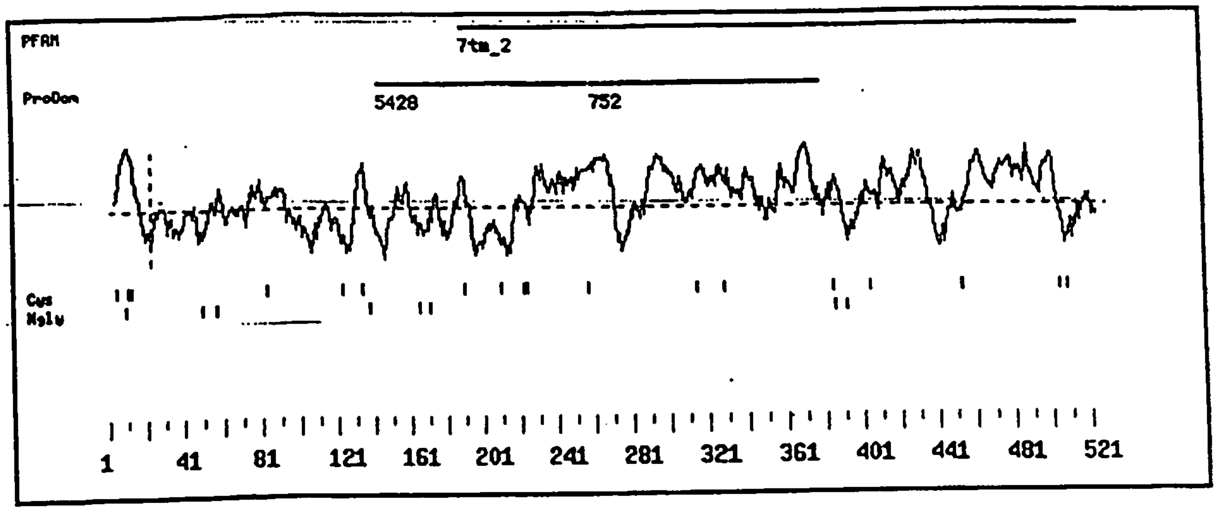
PFAM
7tm_2
ProDom
5428
752
Cys
Nglw
1
41
81
121
161
201
241
281
321
361
401
441
481
521

Figure 25

```
    *->CnrtWDgitC..Wpdt....ppGelVvvpCPkyfygfssdqtdttgn
              +tC  W+ + +++p+G ++   C        + +q   + +
187    -------LTCvfWKEGarkqPWGGWSPEGC-------RTEQ---PSH 216

    vsRnCtedGsWsepppsNrtWrnysaCgeddpeeesekkkkyylvlkiiY
    ++  C+ +    +++             + ++ +   +++ +   l +i
217 SQVLCRCNH--LTYFA------------VLMQLSPALVPAELLAPLTYIS 252

    tvGYSlSLaaLlvAvvILllFRkLhtlwpdnadgalevgapWGAPfqvrr

    +vG S+S++a l+ v++   FRk        + +
253 LVGCSISIVASLITVLLHFHFRKQS-----DSL----------------- 280

    SirCtRNyIHmNLFlSFILrAasvfikdavlksevssdeperLssrcsls
        tR  IHmNL +S +L +++ ++  a   s v+ ++
281 ----TR--IHMNLHASVLLLNIAFLLSPAFAMSPVPGSA----------- 313

    tgqvvvgCkllvvfQfqYcvmtNffWlLvEGlYLhtLLvvtffsErkylw
           C +l ++ ++Y++++  +W+ +EG  L+ LL    +    ++y +
314 -------CTALAAA-LHYALLSCLTWMAIEGFNLYLLLGRVY---NIYIR 352

    wYl....lIGWGvPlVfvtvWaivRllfedtgCWdsnGLAmFPEAKmCiW
     Y+ + +++GWG+P++ v     v++ ++  +C++++    F
353 RYVfklgVLGWGAPALLVLLSLSVKSSVY-GPCTIPV----FDSWENGTG 397

    msdnshlwWIIkgPiLlsilV..........NFflFinIirILvtKLRaa
      n+++ W+ + P++ s+lV + ++ ++  N++++    ++ L + LR+
398 F-QNMSICWV-RSPVVHSVLVmgyggltslfNLVVLAWALWTL-RRLRER 444

    qtgetdqrqYsqYrkLaKSTLlLIPLfGIhyvvFafrPsndarGvlrkik
     + +         + +  L L  L+G++  + +f+++     v+   +
445 ADAPSVR-----ACHDTVTVLGLTVLLGTTWALAFFSFG-----VFLLPQ 484

    lyfelsLgSFQGFfVAvlYCFlNgEVQaEirrrW<-*
    l++   L+S+ GFf  ++  F+ +  ++E   +
485 LFLFTILNSLYGFF--LFLWFCSQRCRSEAEAKA     516
```

Figure 26

```
               10        20        30        40        50        60        70
inputs ATGACGCCGAGCCCCCTGTTGCTGCTCCTGCTGCCGCCGCTGCTGCTGGGGGCCTTCCCGCCGGCCGCCG

       ----------------------------------------------------------------------

               80        90       100       110       120       130       140
inputs CCGCCCGAGGCCCCCAAAGATGGCGGACAAGGTGGTCCCACGGCAGGTGGCCCGGCTGGGCCGCACTGT
                                             ::::
       --------------------------------------CACG----------------------------

              150       160       170       180       190       200       210
inputs GCGGCTGCAGTGCCCAGTGGAGGGGGACCCGCCGCCGCTGACCATGTGGACCAAGGATGGCCGCACCATC

       ----------------------------------------------------------------------

              220       230       240       250       260       270       280
inputs CACAGCGGCTGGAGCCGCTTCCGCGTGCTGCCGCAGGGGCTGAAGGTGAAGCAGGTGGAGCGGGAGGATG

       ----------------------------------------------------------------------

              290       300       310       320       330       340       350
inputs CCGGCGTGTACGTGTGCAAGGCCACCAACGGCTTCGGCAGCCTGAGCGTCAACTACACCCTCGTCGTGCT

       ----------------------------------------------------------------------

              360       370       380       390       400       410       420
inputs GGATGACATTAGCCCAGGGAAGGAGAGCCTGGGGCCCGACAGCTCCTCTGGGGGTCAAGAGGACCCCGCC

       ----------------------------------------------------------------------

              430       440       450       460       470       480       490
inputs AGCCAGCAGTGGGCACGACCGCGCTTCACACAGCCCTCCAAGATGAGGCGCCGGGTGATCGCACGGCCCG
                      ::.:::
       ---------------CGTCCG-------------------------------------------------
                          10

              500       510       520       530       540       550       560
inputs TGGGTAGCTCCGTGCGGCTCAAGTGCGTGGCCAGCGGGCACCCTCGGCCCGACATCACGTGGATGAAGGA

       ----------------------------------------------------------------------

              570       580       590       600       610       620       630
inputs CGACCAGGCCTTGACGCGCCCAGAGGCCGCTGAGCCCAGGAAGAAGAAGTGGACACTGAGCCTGAAGAAC

       ----------------------------------------------------------------------

              640       650       660       670       680       690       700
inputs CTGCGGCCGGAGGACAGCGGCAAATACACCTGCCGCGTGTCGAACCGCGCGGGCGCCATCAACGCCACCT
```

Figure 27A

```
              710       720       730       740       750       760       770
inputs ACAAGGTGGATGTGATCCAGCGGACCCGTTCCAAGCCCGTGCTCACAGGCACGCACCCCGTGAACACGAC

       ----------------------------------------------------------------------

              780       790       800       810       820       830       840
inputs GGTGGACTTCGGGGGGACCACGTCCTTCCAGTGCAAGGTGCGCAGCGACGTGAAGCCGGTGATCCAGTGG

       ----------------------------------------------------------------------

              850       860       870       880       890       900       910
inputs CTGAAGCGCGTGGAGTACGGCGCCGAGGGCCGCCACAACTCCACCATCGATGTGGGCGGCCAGAAGTTTG

       ----------------------------------------------------------------------

              920       930       940       950       960       970       980
inputs TGGTGCTGCCCACGGGTGACGTGTGGTCGCGGCCCGACGGCTCCTACCTCAATAAGCTGCTCATCACCCG
              :::::::::::: ::::::::.::::: :: :::::::::::::: ::::::::::::.: ::
       -------GCCCACGGGTGATGTGTGGTCACGGCCTGATGGCTCCTACCTCAACAAGCTGCTCATCTCTCG
                     20        30        40        50        60        70

              990      1000      1010      1020      1030      1040      1050
inputs TGCCCGCCAGGACGATGCGGGCATGTACATCTGCCTTGGCGCCAACACCATGGGCTACAGCTTCCGCAGC
        ::::::::::: ::::: :::::::::::::::::.:: :: :: :::::::::::::: ::::: :::
       GGCCCGCCAGGATGATGCTGGCATGTACATCTGCCTAGGTGCAAATACCATGGGCTACAGTTTCCGTAGC-
           80        90       100       110       120       130       140

             1060      1070      1080      1090      1100      1110      1120
inputs GCCTTCCTCACCGTGCTGCCAGACCCAAAACCGCCAGGGCCACCTGTGGCCTCCTCGTCCTCGGCCACTA
       ::::::::::: ::. :.:::::::: ::::: ::::::::.:::.:::: :: ::.:: ::. ::::.:
       GCCTTCCTCACTGTATTACCAGACCCCAAACCTCCAGGGCCTCCTATGGCTTCTTCATCGTCATCCACAA
          150       160       170       180       190       200       210

             1130      1140      1150      1160      1170      1180      1190
inputs GCCTGCCGTGGCCCGTGGTCATCGGCATCCCAGCCGGCGCTGTCTTCATCCTGGGCACCCTGCTCCTGTG
       :::::::.::::: ::::: ::::::::::::::: :: :::::::::::::::.:::::  :::: :: ::
       GCCTGCCATGGCCTGTGGTGATCGGCATCCCAGCTGGTGCTGTCTTCATCCTAGGCACTGTGCTGCTCTG
          220       230       240       250       260       270       280

             1200      1210      1220      1230      1240      1250      1260
inputs GCTTTGCCAGGCCCAGAAGAAGCCGTGCACCCCCGCGCCTGCCCCTCCCCTGCCTGGGCACCGCCCGCCG
       ::::::::::.:: :::::::::::.:: .:::: ::. ::.: : :::  ::::::::: :: :: ::.
       GCTTTGCCAGACCAAGAAGAAGCCATGTGCCCCAGCATCTACACTTCCTGTGCCTGGGCATCGTCCCCCA
          290       300       310       320       330       340       350

             1270      1280      1290      1300      1310      1320      1330
inputs GGGACGGCCCGCGACCGCAGCGGAGACAAGGACCTTCCCTCGTTGGCCGCCCTCAGCGCTGGCCCTGGTG
       :::::. :::: :: ::::: ::.:::::::::::: :::::.:::::                     ::
       GGGACATCCCGAGAACGCAGTGGTGACAAGGACCTGCCCTCATTGGC---------------------TG
          360       370       380       390       400

             1340      1350      1360      1370      1380      1390      1400
inputs TGGGGCTGTGTGAGGAGCATGGGTCTCCGGCAGCCCCCCAGCACTTACTGGGCCCAGGCCCAGTTGCTGG
       ::::  :.:::::::::::::::.::  : . .:::::::::::::.: :::: : :.::: ::. ::::::
       TGGGCATATGTGAGGAGCATGGATCCGCCATGGCCCCCCAGCACATCCTGGCCTCTGGCTCAACTGCTGG
          410       420       430       440       450       460       470
```

FIGURE 27B

```
              1410      1420      1430      1440      1450      1460
inputs CCCTAAGTTGTACCCCAAACTCTACACAGACATCCACACACACACACA--CACACAC--TCTCACACACA
       ::: ::: ::::::::::.:: :::::::: .: ::::::::::::::  :::::::   :.:.:::::.
       CCCCAAGCTGTACCCCAAGCTATACACAGATGTGCACACACACACATACACACACCTGCACTCACACG
             480       490       500       510       520       530       540

       1470      1480           1490      1500      1510
inputs CTCACACGT-GGAGGGCAAGGT-C-----CACCAGCACATCCACTATCAGTGC-----------------
       :::.:: :: ::::::::::::: :      :::::::: .::::::::::::::
       CTCTCATGTTGGAGGGCAAGGTTCATCAACACCAGCATGTCCACTATCAGTGCTAAATACAGCGAATCTC
             550       560       570       580       590       600       610

inputs --------------

       CAAGCACTGTGTCC
             620
```

FIGURE 27C

```
970       980       990       1000      1010      1020      1030
GTGCTGCCCACGGGTGACGTGTGGTCGCGGCCCGACGGCTCCTACCTCAATAAGCTGCTCATCACCCGTG
:: : :::::::::::: ::::::::.::::: :: :::::::::::::: ::::::::::::.: :: :
GTCCGGCCCACGGGTGATGTGTGGTCACGGCCTGATGGCTCCTACCTCAACAAGCTGCTCATCTCTCGGG
    10        20        30        40        50        60        70

1040      1050      1060      1070      1080      1090      1100
CCCGCCAGGACGATGCGGGCATGTACATCTGCCTTGGCGCCAACACCATGGGCTACAGCTTCCGCAGCGC
:::::::::: ::::: :::::::::::::::::.:: :: :: :::::::::::::: ::::: :::::
CCCGCCAGGATGATGCTGGCATGTACATCTGCCTAGGTGCAAATACCATGGGCTACAGTTTCCGTAGCGC
    80        90        100       110       120       130       140

1110      1120      1130      1140      1150      1160      1170
CTTCCTCACCGTGCTGCCAGACCCAAAACCGCCAGGGCCACCTGTGGCCTCCTCGTCCTCGGCCACTAGC
:::::::::: ::. :.::::::::: ::::: :::::::.:::.:::: :: ::.:: ::. ::::.:::
CTTCCTCACTGTATTACCAGACCCCAAACCTCCAGGGCCTCCTATGGCTTCTTCATCGTCATCCACAAGC
    150       160       170       180       190       200       210

1180      1190      1200      1210      1220      1230      1240
CTGCCGTGGCCCGTGGTCATCGGCATCCCAGCCGGCGCTGTCTTCATCCTGGGCACCCTGCTCCTGTGGC
:::::.::::: ::::: :::::::::::::: :: ::::::::::::::.:::::  :::: :: ::::
CTGCCATGGCCTGTGGTGATCGGCATCCCAGCTGGTGCTGTCTTCATCCTAGGCACTGTGCTGCTCTGGC
    220       230       240       250       260       270       280

1250      1260      1270      1280      1290      1300      1310
TTTGCCAGGCCCAGAAGAAGCCGTGCACCCCCGCGCCTGCCCCTCCCCTGCCTGGGCACCGCCCGCCGGG
:::::::::.:: :::::::::::.:: .:::: ::. ::.: : :::  :::::::::: :: :: ::.::
TTTGCCAGACCAAGAAGAAGCCATGTGCCCCAGCATCTACACTTCCTGTGCCTGGGCATCGTCCCCCAGG
    290       300       310       320       330       340       350

1320      1330      1340      1350      1360      1370      1380
GACGGCCCGCGACCGCAGCGGAGACAAGGACCTTCCCTCGTTGGCCGCCCTCAGCGCTGGCCCTGGTGTG
:::. :::: :: ::::: ::.:::::::::::: ::::::.:::::                   ::::
GACATCCCGAGAACGCAGTGGTGACAAGGACCTGCCCTCATTGGC---------------------TGTG
    360       370       380       390       400

1390      1400      1410      1420      1430      1440      1450
GGGCTGTGTGAGGAGCATGGGTCTCCGGCAGCCCCCCAGCACTTACTGGGCCCAGGCCCAGTTGCTGGCC
::  :.:::::::::::::::.::   : . .:::::::::::::.: :::: : :.::: ::. ::::::::
GGCATATGTGAGGAGCATGGATCCGCCATGGCCCCCCAGCACATCCTGGCCTCTGGCTCAACTGCTGGCC
    410       420       430       440       450       460       470

1460      1470      1480      1490      1500          1510         1520
CTAAGTTGTACCCCAAACTCTACACAGACATCCACACACACACACA--CACACAC--TCTCACACACACT
: ::: :::::::::::.:: :::::::: .: :::::::::::::::  :::::::   :.:.:::::.::
CCAAGCTGTACCCCAAGCTATACACAGATGTGCACACACACACACATACACACACCTGCACTCACACGCT
    480       490       500       510       520       530       540

  1530      1540                1550      1560      1570      1580
CACACGT-GGAGGGCAAGGT-C-----CACCAGCACATCCACTATCAGTGCTAGACGGCACCGTATCTGC
:.:: :: :::::::::::: :      :::::::: .::::::::::::::::::.:  .:: ::.:::: :
CTCATGTTGGAGGGCAAGGTTCATCAACACCAGCATGTCCACTATCAGTGCTAAA-TACAGCGAATCTCC
    550       560       570       580       590       600       610

1590      1600      1610      1620      1630      1640      1650
AGTGGGCACGGGGGGGCCGGCCAGACAGGCAGACTGGGAGGATGGAGGACGGAGCTGCAGACGAAGGCAG
:.   :::: : :        ::.::  :: ::.:.      .:::.:: :...::..:::  :..::  :
AA---GCACTGTGT------CCTGA--GGTAGGCAT-----TTGGGGGCCAAGGCAACAG--GTTGG--G
         620                630              640       650       660
```

FIGURE 28A

```
1660      1670      1680      1690      1700      1710      1720
GGGACCCATGGCGAGGAGGAATGGCCAGCACCCCAGGCAGTCTGTGTGTGAGGCATAGCCCCTGGACACA
.:.:    : ..:.: :..:.:.:    :: : :  ::: .: :: :.::  .:.::    : :  ..::. .
AGAATTGAGAACAATGGAGGAAG---AGTATCTTAGGGTGCCT-TATGGTGGACA---CTCACAAACTTG
    670       680            690       700        710          720

1730      1740      1750       1760      1770      1780      1790
CACACACAGACACACACACTGCCTGGA-TGCATGTATGCACACACATGCGCGCACACGTGCTCCCTGAAG
  :: : ::: . : . :::.::.:.. ..::     :  :: : .:: .:.:::::: ::   .  : :..
GCCATATAGATGTATGTACTACCAGATGAACAGCCAGCCAGATTCACACACGCACATGTTTAAAC-GTGT
 730       740       750       760       770       780       790

1800      1810       1820      1830      1840      1850      1860
GCACACGTACGCA-CA-CACGCACATGCACAGATATGCCGCCTGGGCACACAGATAAGCTGCCCAAATGC
. ::. ::.:.:: :. :::.::::. : :.:: :..::  :.::. . . .:. ..: :: :  ..:::
AAACGTGTGCACAACTGCACACACAA-C-CTGAGAAACCTTCAGGAGGATTTGTGGTG-TGAC--TTTGC
 800       810       820         830       840       850             860

 1870      1880      1890      1900       1910        1920      1930
ACGCACACGCA-CAGAGACATGCCAGAACATACAAGGACATG-CTGCCTGAACATA--CACACGCACACC
:   ::: : : :...:.:..:  ..:. :..:.   .:::: ::   ::..:::.  :::.  : ::::
AGTGACATGTAGCGATGGCTAGTTGAAGGAATCTCCCTCATGTCTTAGTGGTCATGGCCACTTCCCCACC
      870       880       890       900       910       920       930

     1940          1950      1960      1970      1980      1990
CATGCGCAGATGTG---CTGCCTGGACACACACACACACGGATATGCTGTCTGGACGCACACACGTGC
: ::: ::  ::::   :::::::: :. . . . .:.  ::  :.:::    ::::.    .: :: :..:
CCTGCCCATCTGTGTTCCTGCCTGGCCTTGGTGGTGCTTCCG--TGTGCC--CTGGGTTTTC-CAGGAAC
      940       950       960       970                980       990

2000      2010      2020      2030      2040      2050      2060
AGATATGGTATCCGGACACACACGTGCACAGATATGCTGCCTGGACACACAGATAATGCTGCCTTGACAC
     :.   :.:: :::. . . :.:::      .:::.:::. :   . . .:.:...:: .::     :.:
C---CTATCAACCTGACTGGGGTGAGCA-----GTGCAGCCATGCNTGGAGGTTTGAGCCACC----CTC
    1000      1010      1020                 1030      1040      1050

2070      2080
ACACATGCACGGATATTG
 : :.:::. :......:
CC-CTTGCTAGAGAGAAG
    1060      1070
```

FIGURE 28B

```
               10        20        30        40        50        60        70
inputs MTPSPLLLLLLPPLLLGAFPPAAAARGPPKMADKVVPRQVARLGRTVRLQCPVEGDPPPLTMWTKDGRTI

       ----------------------------------------------------------------------

               80        90       100       110       120       130       140
inputs HSGWSRFRVLPQGLKVKQVEREDAGVYVCKATNGFGSLSVNYTLVVLDDISPGKESLGPDSSSGGQEDPA

       ----------------------------------------------------------------------

              150       160       170       180       190       200       210
inputs SQQWARPRFTQPSKMRRRVIARPVGSSVRLKCVASGHPRPDITWMKDDQALTRPEAAEPRKKKWTLSLKN

       ----------------------------------------------------------------------

              220       230       240       250       260       270       280
inputs LRPEDSGKYTCRVSNRAGAINATYKVDVIQRTRSKPVLTGTHPVNTTVDFGGTTSFQCKVRSDVKPVIQW
                                                                 .::
       ----------------------------------------------------------RVR---------

              290       300       310       320       330       340       350
inputs LKRVEYGAEGRHNSTIDVGGQKFVVLPTGDVWSRPDGSYLNKLLITRARQDDAGMYICLGANTMGYSFRS
                                 :::::::::::::::::::.::::::::::::::::::::::::
       --------------------------PTGDVWSRPDGSYLNKLLISRARQDDAGMYICLGANTMGYSFRS
                                      10        20        30        40

              360       370       380       390       400       410       420
inputs AFLTVLPDPKPPGPPVASSSSATSLPWPVVIGIPAGAVFILGTLLLWLCQAQKKPCTPAPAPPLPGHRPP
       :::::::::::::::.:::::.:::::::::::::::::::::.::::::..::::.::.. :.::::::
       AFLTVLPDPKPPGPPMASSSSSTSLPWPVVIGIPAGAVFILGTVLLWLCQTKKKPCAPASTLPVPGHRPP
        50        60        70        80        90      100       110

              430       440       450       460       470       480
inputs GTARDRSGDKDLPSLAALSAGPGVGLCEEHGSPAAPQHLLGPGPVAGPKLYPKLYTDIHTHTHTHSHTH-
       ::.:.:::::::::::       ::.::::::. ::::.:..:..::::::::::::.:::::::. :: 
       GTSRERSGDKDLPSLA-------VGICEEHGSAMAPQHILASGSTAGPKLYPKLYTDVHTHTHTHTCTHT
       120       130              140       150       160       170       180

                  490       500
inputs -------------SHVEGKVHQHIHYQC
                    : . .:  .
       LSCWRARFINTSMSTISAKYSESPSTVS
              190       200
```

FIGURE 29

```
inputs GT--------------------------------------------------------------------
       .:
       ATGTCACCGCCTCTGTGTCCCCTCCTTCTCCTGGCTGTGGGCCTGCGGCTGGCTGGAACTCTCAACCCCA
               10        20        30        40        50        60        70

inputs ----------------------------------------------------------------------

       GTGATCCCAATACCTGCAGCTTCTGGGAAAGCTTCACTACCACCACCAAGGAGTCCCACTCCCGCCCCTT
               80        90       100       110       120       130       140

inputs ----------------------------------------------------------------------

       CAGCCTGCTCCCCTCAGAGCCCTGCGAGCGGCCCTGGGAGGGCCCCCATACTTGCCCCAGCCCACAAACT
              150       160       170       180       190       200       210

inputs ----------------------------------------------------------------------

       CAGAGGAAACTCCTGGCTTCTAGGGATTCATTCTGCATGGTCTGTGTCGGGGCTGGAGTGCAGTGGCGAG
              220       230       240       250       260       270       280

inputs ----------------------------------------------------------------------

       ATCGTAGTGCACTGCAACCTCAAACAGGGAATGCGCTTTCTATGCGGCCTCAGCCGAGAGTGTTGAGTGG
              290       300       310       320       330       340       350

inputs ----------------------------------------------------------------------

       TGCCCCTTCCCTGGCCTCCCCTGGCCACACTGTGGTGGTGAAGACGGACCACCGCCAGCGCCTGCAGTGC
              360       370       380       390       400       410       420

inputs ----------------------------------------------------------------------

       TGCCATGGCTTCTATGAGAGCAGGGGGTTCTGTGTCCCGCTCTGTGCCCAGGAGTGTGTCCATGGCCGTT
              430       440       450       460       470       480       490

inputs ----------------------------------------------------------------------

       GTGTGGCACCCAATCAGTGCCAATGTGTGCCAGGCTGGCGGGGCGACGACTGTTCCAGTGCCCCGAACTG
              500       510       520       530       540       550       560

inputs ----------------------------------------------------------------------

       CCTTCAGCCCTGTACCCCTGGCTACTATGGCCCTGCCTGCCAGTTCCGCTGCCAGTGCCATGGGGCACCC
              570       580       590       600       610       620       630

inputs ----------------------------------------------------------------------

       TGCGATCCCCAGACTGGAGCCTGCTTCTGCCCCGCAGAGAGAACTGGGCCCAGCTGTGACGTGTCCTGTT
              640       650       660       670       680       690       700
```

Figure 30A

```
inputs ----------------------------------------------------------------------

       CCCAGGGCACTTCTGGCTTCTTCTGCCCCAGCACCCATCCTTGCCAAAATGGAGGTGTCTTCCAAACCCC
               710       720       730       740       750       760       770

inputs ----------------------------------------------------------------------

       ACAGGGCTCCTGCAGCTGCCCCCCTGGCTGGATGGGCACCATCTGCTCCCTGCCCTGCCCAGAGGGCTTT
               780       790       800       810       820       830     . 840

inputs ----------------------------------------------------------------------

       CACGGACCCAACTGCTCCCAGGAATGTCGCTGCCACAACGGCGGCCTCTGTGACCGATTCACTGGGCAGT
               850       860       870       880       890       900       910

inputs ----------------------------------------------------------------------

       GCCGCTGCGCTCCGGGTTACACTGGGGATCGGTGCCGGGAGGAGTGCCCGGTGGGCCGCTTTGGGCAGGA
               920       930       940       950       960       970       980

inputs ----------------------------------------------------------------------

       CTGTGCTGAGACGTGCGACTGCGCCCCGGAGGCCCGTTGCTTCCCGGCCAACGGCGCATGTCTGTGCGAA
               990      1000      1010      1020      1030      1040      1050

inputs ----------------------------------------------------------------------

       CACGGCTTCACTGGGGACCGCTGCACGGATCGCCTCTGCCCCGACGGCTTCTACGGTCTCAGCTGCCAGG
              1060      1070      1080      1090      1100      1110      1120

inputs -------------CGACC----------------------------------------------------
                    :::::
       CCCCCTGCACCTGCGACCGGGAGCACAGCCTCAGCTGCCACCCGATGAACGGGGAGTGCTCCTGCCTGCC
              1130      1140      1150      1160      1170      1180      1190

                                                     10
inputs ------------------------------------------CACGC-----------------------
                                                 :::::
       GGGCTGGGCGGGCCTCCACTGCAACGAGAGCTGCCCGCAGGACACGCATGGGCCAGGGTGCCAGGAGCAC
              1200      1210      1220      1230      1240      1250      1260

inputs ----------------------------------------------------------------------

       TGTCTCTGCCTGCACGGTGGCGTCTGCCAGGCTACCAGCGGCCTCTGTCAGTGCGCGCCGGGTTACACGG
              1270      1280      1290      1300      1310      1320      1330

inputs ----------------------------------------------------------------------

       GCCCTCACTGTGCTAGTCTTTGTCCTCCTGACACCTACGGTGTCAACTGTTCTGCACGCTGCTCATGTGA
              1340      1350      1360      1370      1380      1390      1400
```

FIGURE 30B

```
inputs ----------------------------------------------------------------------

       AAATGCCATCGCCTGCTCACCCATCGACGGCGAGTGCGTCTGCAAGGAAGGTTGGCAGCGTGGTAACTGC
             1410      1420      1430      1440      1450      1460      1470

inputs ----------------------------------------------------------------------

       TCTGTGCCCTGCCCACCCGGAACCTGGGGCTTCAGTTGCAATGCCAGCTGCCAGTGTGCCCATGAGGCAG
             1480      1490      1500      1510      1520      1530      1540

inputs ---------------------------------------------------------------------G
                                                                             :
       TCTGCAGCCCCCAAACTGGAGCCTGTACCTGCACCCCTGGGTGGCATGGGGCCCACTGCCAGCTGCCCTG
             1550      1560      1570      1580      1590      1600      1610

                                                                          20
inputs TCCG----------------------------------------------------------GTGACCCT
       ::::                                                          ::::::::
       TCCGAAGGGGCAGTTTGGAGAAGGTTGTGCCAGTCGCTGTGACTGTGACCACTCTGATGGCTGTGACCCT
             1620      1630      1640      1650      1660      1670      1680

               30        40        50        60        70        80        90
inputs GTTCATGGACAGTGCCGATGTCAGGCTGGTTGGATGGGCACACGCTGCCACCTGCCTTGCCCGGAGGGCT
       ::::::::::. :: :..:: :::::::: :::::::: .: :::::::::::: : ::::: :::::::
       GTTCATGGACGCTGTCAGTGCCAGGCTGGCTGGATGGGTGCCCGCTGCCACCTGTCCTGCCCTGAGGGCT
             1690      1700      1710      1720      1730      1740      1750

              100       110       120       130       140       150       160
inputs TTTGGGGAGCCAACTGCAGTAACACCTGTACCTGCAAGAATGGTGGTACCTGTGTGTCTGAGAATGGCAA
       :.::::::: :::::: :: :::::::: ::::::::::::: :: :::::: :  :::::::::::::
       TATGGGGAGTCAACTGTAGCAACACCTGCACCTGCAAGAATGGGGGCACCTGTCTCCCTGAGAATGGCAA
             1760      1770      1780      1790      1800      1810      1820

              170       180       190       200       210       220       230
inputs CTGCGTGTGCGCACCAGGGTTCCGAGGCCCCTCCTGCCAGAGGCCCTGCCCGCCTGGTCGCTATGGCAAA
       ::::::::: ::::: ::.:::::.:::::::::::::::::::. :::: : :::::: ::::::::::::
       CTGCGTGTGTGCACCCGGATTCCGGGGCCCCTCCTGCCAGAGATCCTGTCAGCCTGGCCGCTATGGCAAA
             1830      1840      1850      1860      1870      1880      1890

              240       250       260       270       280       290       300
inputs CGCTGTGTGCAATGCAAGTGTAACAACAACCATTCTTCCTGCCACCCATCGGACGGGACCTGCTCCTGCC
       ::::::::::  ::::::::    :.  ::::: :: : ::::::::: :::.:::::::::::::: :::::
       CGCTGTGTGCCCTGCAAGTG---CGCTAACCACTCCTTCTGCCACCCCTCGAACGGGACCTGCTACTGCC
             1900      1910           1920      1930      1940      1950

              310       320       330       340       350       360       370
inputs TGGCGGGCTGGACAGGCCCTGACTGCTCCGAGGCATGTCCCCCAGGCCACTGGGGACTCAAATGCTCCCA
       :::: ::::::::::::::: ::::::::: :: :::: :: ::::: ::::::::: . :: ::  ::::
       TGGCTGGCTGGACAGGCCCCGACTGCTCCCAGCCATGCCCTCCAGGACACTGGGGAGAAAACTGTGCCCA
         1960      1970      1980      1990      2000      2010      2020

              380       390       400       410       420       430       440
inputs ACTCTGCCAGTGTCATCATGGTGGGACCTGCCACCCCCAGGATGGGAGCTGTATCTGCACGCCAGGCTGG
       .  ::::::.::::: ::::::::::::::::::: ::::::::::::::::::::::::: : : :::::::
       GACCTGCCAATGTCACCATGGTGGGACCTGCCATCCCCAGGATGGGAGCTGTATCTGCCCCCTAGGCTGG
         2030      2040      2050      2060      2070      2080      2090
```

FIGURE 30C

```
          450       460       470       480       490       500       510
inputs ACTGGACCCAACTGCTTGGAAGGCTGCCCACCAAGAATGTTTGGTGTCAACTGCTCCCAGCTATGTCAGT
       ::::::: : :::::::.:::::::::::.: ..:.: .::::::: ::::::::::::: ::: ::::
       ACTGGACACCACTGCTTAGAAGGCTGCCCTCTGGGGACATTTGGTGCTAACTGCTCCCAGCCATGCCAGT
     2100      2110      2120      2130      2140      2150      2160

          520       530       540       550       560       570       580
inputs GTGATCTCGGAGAGATGTGCCACCCAGAGACTGGGGCTTGTGTCTGTCCCCCAGGACACAGTGGTGCAGA
       :::.::  :::::.:.:::::::::::::::::::: ::::: ::::::::::.:::::::::::
       GTGGTCCTGGAGAAAAGTGCCACCCAGAGACTGGGGCCTGTGTATGTCCCCCAGGGCACAGTGGTGCACC
     2170      2180      2190      2200      2210      2220      2230

          590       600       610       620       630       640       650
inputs CTGCAAAATGGGAAGCCAGGAGTCCTTCACCATAATGCCCACCTCTCCCGTGACCCATAACTCACTGGGT
        :::::..:: :::: ::::::: :::: :: .:.::::: :::.:::: :::..:  :::::::.::::::
       TTGCAGGATTGGAATCCAGGAGCCCTTTACTGTGATGCCGACCACTCCAGTAGCGTATAACTCGCTGGGT
     2240      2250      2260      2270      2280      2290      2300

          660       670       680       690       700       710       720
inputs GCAGTGATTGGCATTGCAGTACTGGGAACCCTCGTGGTGGCCCTGATAGCACTGTTCATTGGCTACCGCC
       :::::::::::::::::::::.:::::..:::: :::::.::::::.:::::::::::::::::: :: :
       GCAGTGATTGGCATTGCAGTGCTGGGGTCCCTTGTGGTAGCCCTGGTGGCACTGTTCATTGGCTATCGGC
     2310      2320      2330      2340      2350      2360      2370

          730       740       750       760       770       780       790
inputs AGTGGCAAAAAGGGCAAGGAACATGAGCACTTGGCAGTGGCTTACAGCACTGGGCGGCTGGATGGCTCTGA
       : ::::::::.::::::::.::  : ::: ::::.:::::::::::::   ::::: ::::: ::::: ::
       ACTGGCAAAAAGGCAAGGAGCACCACCACCTGGCTGTGGCTTACAGCAGCGGGCGCCTGGACGGCTCCGA
     2380      2390      2400      2410      2420      2430      2440

          800       810       820       830       840       850       860
inputs TTACGTCATGCCAGATGTCTCTCCGAGCTATAGTCACTACTACTCCAACCCCAGCTACCACACACTGTCT
       :: ::::::::::::::: :::::::::: ::::::::::::::::::::::::::::::::: :::::
       GTATGTCATGCCAGATGTCCCTCCGAGCTACAGTCACTACTACTCCAACCCCAGCTACCACACCCTGTCG
     2450      2460      2470      2480      2490      2500      2510

          870       880       890       900       910       920       930
inputs CAGTGTTCTCCTAACCCCCCGCCCCCTAACAAGGTCCCAGGCAGTCAGCTCTTTGTCAGCTCTCAGGCCC
       ::::: :: ::.::::::::::::::::::::::::: ::::::    : :::::::: ::::   :::. ::
       CAGTGCTCCCCAAACCCCCCACCCCCTAACAAGGTTCCAGGC---CCGCTCTTTGCCAGCCTGCAGAACC
     2520      2530      2540      2550           2560      2570      2580

          940       950       960       970       980       990      1000
inputs CTGAGCGGCCAAGCAGAGCCCACGGGCGTGAGAACCATACCACACTGCCCGCTGACTGGAAGCACCGCCG
       ::::::::::::.: .:.::::: ::::.::: ::::: ::::: ::::: :::::::::::::::::::
       CTGAGCGGCCAGGTGGGGCCCAAGGGCATGATAACCACACCACCCTGCCTGCTGACTGGAAGCACCGCCG
     2590      2600      2610      2620      2630      2640      2650

         1010                     1020      1030      1040      1050      1060
inputs GGAGCCCCAT------------GACAGAGGCGCCAGCCACCTGGACCGAAGCTATAGCTGTAGCTATAGC
       :::::::: :             :::::.:: . :::::.:::::::::::::: :::.:::::: :::
       GGAGCCCCCTCCAGGGCCTCTGGACAGGGGGAGCAGCCGCCTGGACCGAAGCTACAGCTATAGCTACAGC
     2660      2670      2680      2690      2700      2710      2720

         1070      1080      1090      1100      1110      1120      1130
inputs CACAGGAATGGCCCAGGACCATTCTGTCATAAAGGTCCCATCTCTGAAGAGGGACTAGGGGCAAGCGTTA
             ::::::::::: :::::::.  ::::::: : ::::::::::::::..:: ::::: :: :: .
       ------AATGGCCCAGGCCCATTCTACGATAAAGGGCTCATCTCTGAAGAGGAGCTCGGGGCCAGTGTGG
              2730      2740      2750      2760      2770      2780
```

FIGURE 30D

```
        1140      1150      1160      1170      1180      1190      1200
inputs TGTCCCTGAGCAGTGAGAACCCCTATGCTACCATCCGAGACCTGCCCAGCCTGCCTGGGGAACCCCGAGA
       :::::::::::::::::: ::::: ::::::::.::::::::::: ::::.::::. ::::::.::
       CTTCCCTGAGCAGTGAGAACCCATATGCCACCATCCGGGACCTGCCCAGCTTGCCAGGGGGCCCCCGGGA
     2790      2800      2810      2820      2830      2840      2850

        1210      1220      1230      1240      1250      1260      1270
inputs AAGTGGCTATGTGGAGATGAAAGGACCTCCATCAGTGTCCCCTCCCAGGCAGTCTCTTCATCTCCGGGAC
       .:: .:::: .:::::::::::::: ::::: :::: .:: : ::::::::: ::: ::: : :::::
       GAGCAGCTACATGGAGATGAAAGGCCCTCCCTCAGGATCTGCCCCCAGGCAGCCTCCTCAGTTTTGGGAC
     2860      2870      2880      2890      2900      2910      2920

           1280      1290      1300      1310      1320      1330      1340
inputs AGGCAG---CAGCGGCAACTGCAGCCACAGAGGGACAGCGGCACCTATGAGCAGCCCAGCCCCTTGAGCC
       :: :::   :.:::::::: ::::::::::::.::::: :::::::: :::::::::::::: ::: ::
       AGCCAGAGGCGGCGGCAACCCCAGCCACAGAGAGACAGTGGCACCTACGAGCAGCCCAGCCCCCTGATCC
     2930      2940      2950      2960      2970      2980      2990

           1350      1360      1370      1380      1390      1400      1410
inputs ATAATGAAGAGTCTTTGGGCTCCACGCCCCCGCTTCCTCCAGGCCTGCCTCCTGGTCACTACGACTCCCC
       ::.:  .::: ::: ::::::::  :::::: :: :::::.:::::.:: :: :: ::::: ::::: ::
       ATGACCGAGACTCTGTGGGCTCCCAGCCCCCTCTGCCTCCGGGCCTACCCCCCGGCCACTATGACTCACC
     3000      3010      3020      3030      3040      3050      3060

           1420      1430      1440      1450      1460      1470      1480
inputs CAAGAACAGCCATATCCCTGGACACTATGACTTGCCTCCAGTACGGCATCCTCCATCCCCTCCATCCCGG
       :::::::::::: :::::::::: :::::::::::::::::::::::: ::::: ::::::   ::.
       CAAGAACAGCCACATCCCTGGACATTATGACTTGCCTCCAGTACGGCATCCCCGATCAGCTCCAGTTGGA
     3070      3080      3090      3100      3110      3120      3130

           1490
inputs CGCCAGGACCGC
       :::::::::::
       CGCCAGGACCGT
     3140      3150
```

FIGURE 30E

```
 1890      1900      1910      1920      1930      1940      1950
GACCACTCTGATGGCTGTGACCCTGTTCATGGACGCTGTCAGTGCCAGGCTGGCTGGATGGGTGCCCGCT
:::: :.: :     : :::::::::::::::::. :: :..:: :::::::: :::::::: .: ::::
GACC-CAC-GCGTCCGGTGACCCTGTTCATGGACAGTGCCGATGTCAGGCTGGTTGGATGGGCACACGCT
      10        20        30        40        50        60        70

 1960      1970      1980      1990      2000      2010      2020
GCCACCTGTCCTGCCCTGAGGGCTTATGGGGAGTCAACTGTAGCAACACCTGCACCTGCAAGAATGGGGG
:::::::: : ::::: ::::::::.::::::: :::::: :: ::::::::: ::::::::::::: ::
GCCACCTGCCTTGCCCGGAGGGCTTTGGGGAGCCAACTGCAGTAACACCTGTACCTGCAAGAATGGTGG
       80        90       100       110       120       130       140

 2030      2040      2050      2060      2070      2080      2090
CACCTGTCTCCCTGAGAATGGCAACTGCGTGTGTGCACCCGGATTCCGGGGCCCCTCCTGCCAGAGATCC
:::::: :  ::::::::::::::::::::: ::::: ::.:::::.::::::::::::::::::. ::
TACCTGTGTGTCTGAGAATGGCAACTGCGTGTGCGCACCAGGGTTCCGAGGCCCCTCCTGCCAGAGGCCC
      150       160       170       180       190       200       210

 2100      2110      2120      2130       2140      2150      2160
TGTCAGCCTGGCCGCTATGGCAAACGCTGTGTGCCCTGCAAGTG---CGCTAACCACTCCTTCTGCCACC
:: : :::::: ::::::::::::::::::::  ::::::::    :.  ::::: :: : ::::::::
TGCCCGCCTGGTCGCTATGGCAAACGCTGTGTGCAATGCAAGTGTAACAACAACCATTCTTCCTGCCACC
      220       230       240       250       260       270       280

  2170      2180      2190      2200      2210      2220      2230
CCTCGAACGGGACCTGCTACTGCCTGGCTGGCTGGACAGGCCCCGACTGCTCCCAGCCATGCCCTCCAGG
: :::.:::::::::::: ::::::::: ::::::::::::: ::::::::: :: :::: :: :::::
CATCGGACGGGACCTGCTCCTGCCTGGCGGGCTGGACAGGCCCTGACTGCTCCGAGGCATGTCCCCCAGG
      290       300       310       320       330       340       350

  2240      2250      2260      2270      2280      2290      2300
ACACTGGGGAGAAAACTGTGCCCAGACCTGCCAATGTCACCATGGTGGGACCTGCCATCCCCAGGATGGG
:::::::: . :: ::  ::::.  ::::::.::::: :::::::::::::::: :::::::::::
CCACTGGGGACTCAAATGCTCCCAACTCTGCCAGTGTCATCATGGTGGGACCTGCCACCCCCAGGATGGG
      360       370       380       390       400       410       420

  2310      2320      2330      2340      2350      2360      2370
AGCTGTATCTGCCCCCTAGGCTGGACTGGACACCACTGCTTAGAAGGCTGCCCTCTGGGGACATTTGGTG
:::::::::::: : : ::::::::::::: : ::::::.:::::::::::::.: ..:.: .:::::::
AGCTGTATCTGCACGCCAGGCTGGACTGGACCCAACTGCTTGGAAGGCTGCCCACCAAGAATGTTTGGTG
      430       440       450       460       470       480       490

  2380      2390      2400      2410      2420      2430      2440
CTAACTGCTCCCAGCCATGCCAGTGTGGTCCTGGAGAAAAGTGCCACCCAGAGACTGGGGCCTGTGTATG
 :::::::::::: ::: :::::::.::  ::::.:.::::::::::::::::::::: ::::: ::
TCAACTGCTCCCAGCTATGTCAGTGTGATCTCGGAGAGATGTGCCACCCAGAGACTGGGGCTTGTGTCTG
      500       510       520       530       540       550       560

  2450      2460      2470      2480      2490      2500      2510
TCCCCCAGGGCACAGTGGTGCACCTTGCAGGATTGGAATCCAGGAGCCCTTTACTGTGATGCCGACCACT
:::::::::.::::::::::::    ::::..:: :::: ::::::: :::: :: .:.::::: :::.::
TCCCCCAGGACACAGTGGTGCAGACTGCAAAATGGGAAGCCAGGAGTCCTTCACCATAATGCCCACCTCT
      570       580       590       600       610       620       630

  2520      2530      2540      2550      2560      2570      2580
CCAGTAGCGTATAACTCGCTGGGTGCAGTGATTGGCATTGCAGTGCTGGGGTCCCTTGTGGTAGCCCTGG
:: ::..:  ::::::::.:::::::::::::::::::::::::.::::::..:::: ::::::.::::::.
CCCGTGACCCATAACTCACTGGGTGCAGTGATTGGCATTGCAGTACTGGGAACCCTCGTGGTGGCCCTGA
      640       650       660       670       680       690       700
```

FIGURE 31A

```
     2590      2600      2610      2620      2630      2640      2650
TGGCACTGTTCATTGGCTATCGGCACTGGCAAAAAGGCAAGGAGCACCACCACCTGGCTGTGGCTTACAG
:.:::::::::::::::::: :: :: :::::::.::::::::.::  : ::: ::::.:::::::::::
TAGCACTGTTCATTGGCTACCGCCAGTGGCAAAAGGGCAAGGAACATGAGCACTTGGCAGTGGCTTACAG
      710       720       730       740       750       760       770

     2660      2670      2680      2690      2700      2710      2720
CAGCGGGCGCCTGGACGGCTCCGAGTATGTCATGCCAGATGTCCCTCCGAGCTACAGTCACTACTACTCC
::  ::::: ::::: ::::: :: :: ::::::::::::::: :::::::::: ::::::::::::::
CACTGGGCGGCTGGATGGCTCTGATTACGTCATGCCAGATGTCTCTCCGAGCTATAGTCACTACTACTCC
      780       790       800       810       820       830       840

     2730      2740      2750      2760      2770      2780
AACCCCAGCTACCACACCCTGTCGCAGTGCTCCCCAAACCCCCCACCCCCTAACAAGGTTCCAGGC---C
:::::::::::::::: ::::: ::::: :: ::.::::::::.:::::::::::::: ::::::    :
AACCCCAGCTACCACACACTGTCTCAGTGTTCTCCTAACCCCCCGCCCCCTAACAAGGTCCCAGGCAGTC
      850       860       870       880       890       900       910

2790      2800      2810      2820      2830      2840      2850
CGCTCTTTGCCAGCCTGCAGAACCCTGAGCGGCCAGGTGGGGCCCAAGGGCATGATAACCACACCACCCT
:::::::: ::::   :::. :::::::::::::.: .:.:::::: ::::.::: ::::: ::::: ::
AGCTCTTTGTCAGCTCTCAGGCCCCTGAGCGGCCAAGCAGAGCCCACGGGCGTGAGAACCATACCACACT
      920       930       940       950       960       970       980

2860      2870      2880      2890      2900      2910      2920
GCCTGCTGACTGGAAGCACCGCCGGGAGCCCCTCCAGGGCCTCTGGACAGGGGGAGCAGCCGCCTGGAC
::: ::::::::::::::::::::::::::: :            :::::.:: . :::::.:::::::
GCCCGCTGACTGGAAGCACCGCCGGGAGCCCCAT-------------GACAGAGGCGCCAGCCACCTGGAC
      990      1000      1010                1020      1030

2930      2940      2950            2960      2970      2980      2990
CGAAGCTACAGCTATAGCTACAGC------AATGGCCCAGGCCCATTCTACGATAAAGGGCTCATCTCTG
:::::::: ::::.:::::: :::      ::::::::::: :::::::.   :::::: : ::::::::
CGAAGCTATAGCTGTAGCTATAGCCACAGGAATGGCCCAGGACCATTCTGTCATAAAGGTCCCATCTCTG
1040      1050      1060      1070      1080      1090      1100

    3000      3010      3020      3030      3040      3050      3060
AAGAGGAGCTCGGGGCCAGTGTGGCTTCCCTGAGCAGTGAGAACCCATATGCCACCATCCGGGACCTGCC
::::::..:: ::::: :: :: .  :::::::::::::::::::: ::::: ::::::::.::::::::
AAGAGGGACTAGGGGCAAGCGTTATGTCCCTGAGCAGTGAGAACCCCTATGCTACCATCCGAGACCTGCC
1110      1120      1130      1140      1150      1160      1170

    3070      3080      3090      3100      3110      3120      3130
CAGCTTGCCAGGGGGGCCCCCGGGAGAGCAGCTACATGGAGATGAAAGGCCCTCCCTCAGGATCTGCCCCC
:::: ::::.::::. ::::::.::.:: .:::: .:::::::::::::: ::::: :::: .::  : :::
CAGCCTGCCTGGGGAACCCCGAGAAAGTGGCTATGTGGAGATGAAAGGACCTCCATCAGTGTCCCCTCCC
1180      1190      1200      1210      1220      1230      1240

    3140      3150      3160      3170      3180      3190      3200
AGGCAGCCTCCTCAGTTTTGGGACAGCCAGAGGCGGCGGCAACCCCAGCCACAGAGAGACAGTGGCACCT
:::::: ::: :::  :  ::::::: :.:     :.:::::::: :::::::::::.::::: :::::::
AGGCAGTCTCTTCATCTCCGGGACAGGCAG---CAGCGGCAACTGCAGCCACAGAGGGACAGCGGCACCT
1250      1260      1270           1280      1290      1300      1310

    3210      3220      3230      3240      3250      3260      3270
ACGAGCAGCCCAGCCCCCTGATCCATGACCGAGACTCTGTGGGCTCCCAGCCCCCTCTGCCTCCGGGCCT
: :::::::::::::: ::: ::::.:  .::: ::: ::::::::  :::::: :: :::::.:::::
ATGAGCAGCCCAGCCCCTTGAGCCATAATGAAGAGTCTTTGGGCTCCACGCCCCCGCTTCCTCCAGGCCT
    1320      1330      1340      1350      1360      1370      1380
```

FIGURE 31B

```
   3280      3290      3300      3310      3320      3330      3340
ACCCCCCGGCCACTATGACTCACCCAAGAACAGCCACATCCCTGGACATTATGACTTGCCTCCAGTACGG
.:: :: :: ::::: ::::: ::::::::::::::: ::::::::::: ::::::::::::::::::::::
GCCTCCTGGTCACTACGACTCCCCCAAGAACAGCCATATCCCTGGACACTATGACTTGCCTCCAGTACGG
1390      1400      1410      1420      1430      1440      1450

   3350      3360      3370      3380      3390      3400      3410
CATCCCCCATCACCTCCACTTCGACGCCAGGACCGTTGAGGAGCCAGGATGGTATGGCAGAGGCCAGCAC
::::: ::::: ::::::   ::.::::::::::: :::.:::::.: ::::::::    :..::
CATCCTCCATCCCCTCCATCCCGGCGCCAGGACCGCTGAAGAGCCGGCATGGTATG---GGAGC------
1460      1470      1480      1490      1500      1510

   3420      3430      3440      3450      3460      3470      3480
ACCTGGCTGTTGCTGCTCAAGGCTGGGGACAGAGCCTAGTGTACCCCTGCCAGGAGCAGGGAGTGGACCG
                               :.::::: ::::::  ::::::::::::::: ::::::.
-------------------------------GTGCCTA-TGTACCT-TGCCAGGAGCAGGGACTGGACCA
                               1520      1530      1540      1550

   3490      3500      3510      3520      3530      3540      3550
GCAGGCTGTGAACATGAACAACGCTTAACAGAGCAAGTGATGG-GAGCCTTGTTCCTGGG-TTCTACCAT
:::::: . ::::: .::::   :::.. ..:: .:. ...:. :..: :: :::: : ::: :::..
GCAGGCCACGAACAGAAACA---CTTGGTGAAGTGAACAGAGACGGACTGTGGCCCTGTGCTTCCACCGA
  1560      1570           1580      1590      1600      1610      1620

     3560      3570      3580      3590      3600      3610
GGGAGACGCTGATCAGCAGGATGCCTGGCTCCCTTTCCCAACCCACTGCTCCCAAGGCCTCCAGGGC---
:::::::.::..: ..::...:: ::..: : :::: :::::::::::::  :::: ::   .::.:
GGGAGACACTAGTTGACAAAGTGTCTAACCCTCTTTTCCAACCCACTGCT--CAAGTCCCTGTGGACATA
    1630      1640      1650      1660      1670      1680

3620      3630      3640      3650      3660      3670      3680
--CCTGTGTACATAAACTGGTGGGTTGGAAGTTGCTGGGTAAC-TCTGATTTCAGACATGCGTGTGGGGT
  :  ::: .:: ::. : :: : . ....:: . :  . :.: . : .::: ..: .:. :::: ::::
AGCTGGTGGGCAGAATGTTGTTGTACAAGTGTGATTTTAGATCGATTTTTTTTTAAAGTATGTGTTGGGT
1690      1700      1710      1720      1730      1740      1750

 3690      3700      3710      3720      3730      3740      3750
ACCTTTTCTGTGC--ATGCTCAGCCTGGGCTCTGTGCGTGTGTGTGTTTCTGTGATTTTAGAAGGGTACC
::::::::::::   :::::::: :.::     :::::     ::::: : :   ::. ::::::.::. .:
ACCTTTTCTGTGTGTATGCTCAGGCAGG---CTGTG----TGTGTCTCTAGTTGGCTTTAGAGGGAGTCA
1760      1770      1780      1790      1800      1810      1820

    3760      3770      3780      3790      3800      3810      3820
AG-GCAGGTTCTGTCCTAGGGCACTTACCATTTAGTAGGGAGATGGAACCAACCCAATTAACTCTAGCAA
.: . ::::::::: :::.  ::::::.:::: :..: .:...: ..  :..:: :. :.: .:::: .:
GGTATAGGTTCTG-CCTTCTGCACTTTCCATCTTATCTAGTAGTCAG--CTTCCAAGCTTA-ACTAGTTA
   1830      1840      1850      1860      1870      1880

    3830      3840      3850      3860      3870      3880      3890
TAGCCTCCTAACTGGCCTCCTCCATTGATTCAGTGAACCTTCCAATGCATGGCTCATAATTTCAAAATAC
 ::: :::.       ::. :. ::  :.  :... .:::: ::.           :: :    ::::     :
GAGC-TCCA------CCAGCAGCA--GGCCCTAACTACCTGCCT-------GCCC-----TTCA-----C
1890            1900      1910      1920                1930

    3900      3910      3920      3930      3940      3950      3960
AGGCTGGTTAGTTACTCCCTACCTGAAAGCCTTCATAGGTGCCTCTTTGCTCTTCTGCCAGTATCAAAAC
   : .::.:  : :::: :. ::   ..:: ::: :::.      ::::::     ::     :.: .:
---CCAGTAA--TCCTCCATGTCT--TTGC--TCAGAGGA------TTGCTC-----CC-----CGACTC
        1940      1950           1960                1970
```

- FIGURE 31C

```
   3970      3980      3990      4000      4010      4020      4030
TTTTGAAGGCCTTAAAGGCCCTGCTTTGCCTGGCCCATCTGTCTCTCCAGCCTCACCTTGAACTGTGTTC
TGGTGTTGTCCT-------CCTGGTACGCCTTG----ACGGTC-CTGCAGTCTC-CCT---------TTC
1980             1990      2000           2010      2020

   4040      4050      4060      4070      4080      4090      4100
CTGTCACTGCACGCCAGTCACACCGGCCTCTAGGTCCTCCTGTAGGCCACTCTTCTTTCTGGCACAGGGA
CCGTCT-TGCT--TCATTCTTTC---CCAGAATGAAGGC-TGTCTGCCACCCTACTTCCCAGCCCAGGAA
  2030      2040      2050      2060      2070      2080

   4110      4120      4130      4140      4150      4160      4170
CCTGCACACCTGGAGTGCCCTTCCTCCCCCACTCGCCTGTTCACCCCTGCTTTTCCTTTACACCTCCTCC
TTGGCACATCTAAGTT---CAGCCTTCCTAAGTTACCCGTTGAGTCCTGCTTGCCCTT--CACATATTCC
2090      2100          2110      2120      2130      2140        2150

   4180      4190      4200      4210      4220      4230      4240
TCAGGGAAGTGCCCACCCTCCGTACATCTTTCACAGCCCTGATTGCAGCTGTGTTCACTCACCAGGTACC
ACAGAACA---CCCACC--CC--ACATCT---GCTTC----ATAGCTACTCTCTTCTC-CAC---GTACC
  2160              2170      2180          2190      2200

   4250      4260      4270      4280      4290      4300      4310
TGCAGAAGGCCTACAGGGTGCCAGGCACTTCTTTAATGGGTTCTTTCTTTATGTGATTATTTGATTAATC
CACAGAAGGCAGAAGTGGTACCAGGCAAGA--AGATGGGATTGTTGCATTTTGT--TTTGTTTTTGAGAC
 2210      2220      2230        2240      2250        2260      2270

   4320      4330      4340      4350       4360      4370      4380
TCTGCCTCCCCCACTAGACTGTAAGCTCCCTGAAGGCAAGAATCCTG--TGCTTATGCTCAATATTAGCT
TCTGTCTCACTATGTAGTCCTGGCTGGCCTGGAACTCAAGAGCTCTGCCTGCCTCTGCCTCTTGAGTGCT
     2280      2290      2300      2310      2320      2330      2340

         4390        4400      4410       4420      4430      4440
CTCCCTT--GGCACAGAGT---AGGCACTCAACAAA-TGCTCCCCAAAAGGCTGAGTGGCTGACTGAATT
GGGTTTAACGGCTCAGGGTCACATGCACAGCTCAAGCTGCACTCCGATGTGCT---TTCC--CCTGTTGC
     2350      2360      2370      2380      2390          2400

  4450      4460      4470      4480      4490      4500       4510
AAGTACCAGTGACATGCAGTAACTGCTAAGATAGATGAGCCATCTGTATGCTCTGACAGTTACAG-ACTG
TAGAT----TAGCGT-CTGCCTCCCCCTAG-TGGAGAGGCTGATCGCCAGCTCT--CTGATGCAGGACTC
2410            2420      2430      2440      2450        2460

    4520      4530      4540      4550      4560      4570      4580
AATAAGTTGGAGACT-TCCCTAAAGGGTGGCATTTCCCCAGGGTAACAACGCAGAGCTCAGGTGTGGGAA
TGGTGTTTAGGCTCACTCACTATTGGTTT-CCTTGGCACAGGGTAGTCACTCAATA--AATGTTCCTCTA
2470      2480      2490       2500      2510      2520       2530

    4590      4600      4610
GGTGCCAGGGGCAGGGGTGCAGAGGGGCTGAGGC
AAAGCTGAAAAAAAAAAAAAAAAAGGGCGGCCGC
 2540      2550      2560
```

FIGURE 31D

```
                 10        20        30        40        50        60        70
inputs  MSPPLCPLLLLAVGLRLAGTLNPSDPNTCSFWESFTTTTKESHSRPFSLLPSEPCERPWEGPHTCPSPQT

        ----------------------------------------------------------------------

                 80        90       100       110       120       130       140
inputs  QRKLLASRDSFCMVCVGAGVQWRDRSALQPQTGNALSMRPQPRVLSGAPSLASPGHTVVVKTDHRQRLQC

        ----------------------------------------------------------------------

                150       160       170       180       190       200       210
inputs  CHGFYESRGFCVPLCAQECVHGRCVAPNQCQCVPGWRGDDCSSAPNCLQPCTPGYYGPACQFRCQCHGAP

        ----------------------------------------------------------------------

                220       230       240       250       260       270       280
inputs  CDPQTGACFCPAERTGPSCDVSCSQGTSGFFCPSTHPCQNGGVFQTPQGSCSCPPGWMGTICSLPCPEGF

        ----------------------------------------------------------------------

                290       300       310       320       330       340       350
inputs  HGPNCSQECRCHNGGLCDRFTGQCRCAPGYTGDRCREECPVGRFGQDCAETCDCAPDARCFPANGACLCE

        ----------------------------------------------------------------------

                360       370       380       390       400       410       420
inputs  HGFTGDRCTDRLCPDGFYGLSCQAPCTCDREHSLSCHPMNGECSCLPGWAGLHCNESCPQDTHGPGCQEH
                                                                    .::..:
        ------------------------------------------------------------STHASG----

                430       440       450       460       470       480       490
inputs  CLCLHGGVCQATSGLCQCAPGYTGPHCASLCPPDTYGVNCSARCSCENAIACSPIDGECVCKEGWQRGNC

        ----------------------------------------------------------------------

                500       510       520       530       540       550       560
inputs  SVPCPPGTWGFSCNASCQCAHEAVCSPQTGACTCTPGWHGAHCQLPCPKGQFGEGCASRCDCDHSDGCDP
                                                                            ::
        --------------------------------------------------------------------DP

                570       580       590       600       610       620       630
inputs  VHGRCQCQAGWMGARCHLSCPEGLWGVNCSNTCTCKNGGTCLPENGNCVCAPGFRGPSCQRSCQPGRYGK
        :::.:.:::::::.::::.::::.::.:::::::::::::::..:::::::::::::::::::.:.:::::
        VHGQCRCQAGWMGTRCHLPCPEGFWGANCSNTCTCKNGGTCVSENGNCVCAPGFRGPSCQRPCPPGRYGK
        10        20        30        40        50        60        70
```

Figure 32A

```
                640          650          660          670          680          690
inputs RCVPCKCAN-HSFCHPSNGTCYCLAGWTGPDCSQPCPPGHWGENCAQTCQCHHGGTCHPQDGSCICPLGW
       :::.:::.: :: ::::.::: :::::::::::..::::::: .:.: :::::::::::::::::  ::
       RCVQCKCNNNHSSCHPSDGTCSCLAGWTGPDCSEACPPGHWGLKCSQLCQCHHGGTCHPQDGSCICTPGW
       80           90          100          110          120          130          140

     700          710          720          730          740          750          760
inputs TGHHCLEGCPLGTFGANCSQPCQCGPGEKCHPETGACVCPPGHSGAPCRIGIQEPFTVMPTTPVAYNSLG
       :: .:::::::    ::.:::: :::. ::.:::::::::::::::::: :..: ::.::.:::.::. ::::
       TGPNCLEGCPPRMFGVNCSQLCQCDLGEMCHPETGACVCPPGHSGADCKMGSQESFTIMPTSPVTHNSLG
      150          160          170          180          190          200          210

     770          780          790          800          810          820          830
inputs AVIGIAVLGSLVVALVALFIGYRHWQKGKEHHHLAVAYSSGRLDGSEYVMPDVPPSYSHYYSNPSYHTLS
       :::::::::.:::::.::::::::.:::::::: :::::::.::::::.::::::.::::::::::::::::
       AVIGIAVLGTLVVALIALFIGYRQWQKGKEHEHLAVAYSTGRLDGSDYVMPDVSPSYSHYYSNPSYHTLS
      220          230          240          250          260          270          280

     840          850          860          870          880          890          900
inputs QCSPNPPPPNKVPGP-LFASLQNPERPGGAQGHDNHTTLPADWKHRREPPPGPLDRGSSRLDRSYSYSYS
       :::::::::::::::. ::.: :.::::. :.:..::::::::::::::::     :::.:.:::::: :::
       QCSPNPPPPNKVPGSQLFVSSQAPERPSRAHGRENHTTLPADWKHRREPH----DRGASHLDRSYSCSYS
      290          300          310          320          330                 340          350

         910          920          930          940          950          960          970
inputs --NGPGPFYDKGLISEEELGASVASLSSENPYATIRDLPSLPGGPRESSYMEMKGPPSGSAPRQPPQFWD
         :::::: .:: :::: ::::: ::::::::::::::::::::: ::::.:.:::::::: :.:::. ...:
       HRNGPGPFCHKGPISEEGLGASVMSLSSENPYATIRDLPSLPGEPRESGYVEMKGPPSVSPPRQSLHLRD
         360          370          380          390          400          410          420

         980          990         1000         1010         1020         1030         1040
inputs SQRRRQPQPQRDSGTYEQPSPLIHDRDSVGSQPPLPPGLPPGHYDSPKNSHIPGHYDLPPVRHPPSPPLR
        :.: : :::::::::::::::: :. .:.:: :::::::::::::::::::::::::::::::::::::: :
       RQQR-QLQPQRDSGTYEQPSPLSHNEESLGSTPPLPPGLPPGHYDSPKNSHIPGHYDLPPVRHPPSPPSR
          430          440          450          460          470          480          490

       1050
inputs RQDR
       ::::
       RQDR
```

Figure 32B

GTCCGACCCACGCGTCCGAGCCACACCCTGAAGGTGGTTGGAAGGAGGGAAGGATCTAGGTCCTGAGCACTGGAATTCC 79

CCAGAACAGCATCTGGCTTCCCAGACCCATGCTGGCCACCACTGATGTGTCCTTCCGGCTGCTGGCTGCAGTGCTGTTC 158

TGTTGTTGGGTGCCCTGTGGCAGGCTTGTGCAATGCCACTCTGTCCCCTCCTCCTCCTGGCCCTAGGCCTGCGTCTGGC 237

TGGAACACTCAACTCCAATGATCCCAATGTCTGTACCTTCTGGGAAAGCTTCACCACGACCACTAAGGAGTCCCACCTT 316

CGCCCCTTCAGCCTGCCCCCAGCCGAGTCCTGCGACAGGCCCTGGGAAGACCCCCACACCTGCGCTCAGCCTACGGTTG 395

TCTACCGGACTGTGTACCGTCAGGTGGTGAAGATGGACTCCCGCCCACGCCTGCAGTGCTGTGGGGTTACTACGAGAG 474

CAGTGGAGCCTGTGTCCCACTCTGTGCCCAGGAGTGTGTCCACGGTCGCTGTGTGGCTCCTAATCGGTGCCAGTGTGCA 553

CCAGGCTGGCGGGGTGACGACTGTTCCAGTGAGTGTGCTCCTGGAATGTGGGGACCACAGTGTGACAGGCTCTGCCTCT 632

GTGGCAACAGCAGTTCCTGTGATCCCAGGAGTGGGGTGTGTTTTTGCCCCTCTGGCCTGCAGCCCCCCGACTGCCTTCA 711

GCCTTGCCCCGATGGCCACTATGGTCCTGCCTGCCAGTTTGATTGCCATTGCTATGGGGCATCCTGTGACCCCCGGGAT 790

GGAGCCTGCTTCTGCCCCCCAGGGAGAACAGGACCCAGGGCACTGATGGCTTCTTCTGCCCCAGAACTTATCCTTGCCA 869

M G V I C S 6
AAATGGAGGTGTTCCTCAGGGCTCTCAAGGCTCCTGCAGCTGCCCACCGGGCTGG ATG GGT GTC ATC TGT TCC 942

L P C P E G F H G P N C T Q E C R C H N 26
CTG CCA TGC CCA GAG GGT TTC CAC GGA CCC AAC TGT ACT CAG GAA TGT CGT TGC CAC AAT 1002

G G L C D R F T G Q C H C A P G Y I G D 46
GGT GGC CTT TGT GAC AGG TTT ACT GGG CAG TGC CAC TGT GCT CCT GGC TAT ATC GGG GAT 1062

R C R E E C P V G R F G Q D C A E T C D 66
CGG TGC CGT GAA GAG TGC CCT GTG GGC CGC TTC GGT CAA GAC TGT GCT GAG ACC TGT GAC 1122

C A P G A R C F P A N G A C L C E H G F 86
TGT GCT CCT GGC GCT CGT TGC TTT CCT GCC AAT GGC GCG TGT CTG TGC GAA CAT GGC TTC 1182

T G D R C T E R L C P D G R Y G L S C Q 106
ACA GGC GAC CGC TGC ACT GAG CGA CTC TGT CCA GAT GGC CGC TAT GGT CTG AGC TGC CAA 1242

D P C T C D P E H S L S C H P M H G E C 126
GAT CCC TGC ACC TGC GAC CCA GAA CAC AGT CTC AGC TGC CAC CCA ATG CAC GGC GAG TGC 1302

S C Q P G W A G L H C N E S C P Q D T H 146
TCC TGC CAG CCA GGT TGG GCG GGC CTC CAC TGC AAC GAG AGC TGC CCT CAG GAC ACG CAC 1362

G A G C Q E H C L C L H G G V C L A D S 166
GGA GCC GGT TGC CAG GAG CAC TGC CTC TGT CTG CAC GGC GGT GTT TGC CTC GCC GAC AGC 1422

G L C R C A P G Y T G P H C A N L C P P 186
GGC CTC TGC CGG TGT GCA CCT GGC TAC ACG GGA CCT CAC TGC GCT AAT CTT TGT CCA CCT 1482

N T Y G I N C S S H C S C E N A I A C S 206
AAC ACT TAT GGG ATC AAC TGT TCC TCC CAC TGC TCC TGT GAA AAT GCC ATT GCC TGC TCT 1542

P V D G T C I C K E G W Q R G N C S V P 226
CCT GTC GAC GGC ACG TGC ATC TGC AAG GAA GGT TGG CAG CGT GGT AAC TGC TCT GTG CCC 1602

FIGURE 33A

C P P G T W G F S C N A S C Q C A H E G 246

TGT CCC CCT GGC ACC TGG GGC TTC AGT TGC AAT GCC AGT TGC CAG TGT GCC CAC GAG GGA 1662

V C S P Q T G A C T C T P G W R G V H C 266

GTC TGC AGC CCC CAA ACT GGA GCC TGT ACT TGC ACC CCT GGG TGG CGT GGG GTT CAC TGC 1722

Q L P C P K G Q F G E G C A S V C D C D 286

CAA CTT CCG TGC CCG AAG GGA CAG TTT GGT GAA GGT TGT GCC AGT GTC TGT GAC TGT GAC 1782

H S D G C D P V H G H C R C Q A G W M G 306

CAC TCC GAT GGC TGT GAC CCT GTT CAT GGA CAC TGC CGA TGT CAG GCT GGC TGG ATG GGC 1842

T R C H L P C P E G F W G A N C S N A C 326

ACA CGT TGC CAC CTG CCT TGC CCA GAG GGC TTT TGG GGA GCC AAC TGC AGC AAT GCC TGT 1902

T C K N G G T C V P E N G N C V C A P G 346

ACC TGC AAG AAT GGT GGC ACT TGT GTA CCT GAG AAC GGC AAC TGT GTG TGC GCA CCA GGG 1962

F R G P S C Q R P C P P G R Y G K R C V 366

TTC AGA GGC CCC TCC TGC CAG AGG CCC TGC CCG CCT GGT CGC TAT GGC AAA CGC TGT GTG 2022

P C K C N N H S S C H P S D G T C S C L 386

CCC TGC AAG TGC AAC AAC CAT TCT TCC TGC CAC CCG TCG GAT GGG ACC TGC TCC TGC CTG 2082

A G W T G P D C S E S C P P G H W G L K 406

GCA GGC TGG ACA GGC CCT GAC TGC TCT GAA TCA TGT CCC CCA GGC CAC TGG GGA CTC AAA 2142

C S Q P C Q C H H G A T C H P Q D G S C 426

TGC TCC CAA CCC TGC CAG TGT CAT CAT GGT GCC ACC TGC CAC CCC CAG GAT GGG AGC TGT 2202

V C I P G W T G P N C S E G C P S R M F 446

GTC TGC ATC CCA GGC TGG ACT GGA CCC AAC TGC TCG GAA GGC TGC CCA TCA AGA ATG TTT 2262

G V N C S Q L C Q C D P G E M C H P E T 466

GGT GTC AAC TGC TCC CAG CTA TGT CAG TGT GAT CCT GGA GAG ATG TGC CAC CCA GAG ACT 2322

G A C V C P P G H S G A H C K V G S Q E 486

GGG GCT TGC GTC TGT CCC CCA GGA CAC AGT GGT GCG CAC TGC AAA GTG GGC AGC CAG GAG 2382

S F T I M P T S P V I H N S L G A V I G 506

TCC TTC ACC ATA ATG CCC ACC TCT CCT GTG ATC CAT AAC TCA CTG GGT GCC GTG ATT GGC 2442

I A V L G T L V V A L V A L F I G Y R H 526

ATT GCA GTG CTG GGG ACC CTT GTG GTG GCC CTG GTA GCA CTG TTT ATT GGC TAC CGA CAC 2502

W Q K G K E H E H L A V A Y S T G R L D 546

TGG CAA AAG GGC AAG GAA CAT GAG CAC TTG GCA GTG GCT TAC AGC ACT GGG CGA CTG GAT 2562

G S D Y V M P D V S P S Y S H Y Y S N P 566

GGC TCC GAT TAC GTC ATG CCA GAT GTC TCT CCG AGC TAC AGT CAC TAC TAT TCC AAC CCT 2622

S Y H T L S Q C S P N P P P P N K I P G 586

AGC TAC CAC ACA CTG TCT CAG TGT TCT CCT AAC CCT CCA CCC CCT AAC AAG ATT CCA GGC 2682

S Q L F V S S Q A S E R P N R N H G R D 606

AGT CAG CTG TTT GTC AGC TCC CAG GCA TCT GAG CGG CCA AAC AGA AAC CAT GGG CGA GAT 2742

FIGURE 33B

```
 N   H   A   T   L   P   A   D   W   K   H   R   R   E   S   H   D   R   A   F     626
AAC CAC GCC ACA CTG CCC GCT GAC TGG AAG CAC CGA CGG GAG TCC CAT GAC AGA GCT TTC   2802

 L   R   H   Q   P   P   G   P   K   V   *                                         637
CTC AGG CAC CAG CCA CCT GGA CCG AAG GTA TAG                                       2835

CTGTAGCTATGGCCACAGGAATGGCCCGGGGCCATTCTGTCATAAAGGTCCCATCTCTGAAGAAGGACTAGGGGCAAGC   2914

GTTATGTCCCTGAGCAGTGAGAACCCCTATGCGACCATCCGAGACCTGCCCGGCCTGCCTGGGGAACCCCGAGAAAGCA   2993

GCTATGTGGAGATGAAAGGCCCTCCATCAGTGTCTCCCCCCAGGCAGCCTCTTCATCTCCGGGACAGGCAGCAGCAGCA   3072

ACTGCAGTCTCAGAGAGACAGCGGCACCTATGAGCAGCCCACTCCCTTGAGCCGTAATGAAGAGTCTGTGGGCTCCATG   3151

CCCCCTCTTCCTCCGGGCCTGCCACCCGGCCACTATGACTCGCCCAAAAACAGCCACATCCCTGGACACTATGACTTGC   3230

CTCCAGTACGGCATCCTCCATCACCTCCATCCCGGCGCCAGGACCGCTGAGGAGCCAGCATGGTATGGAGAGTGCCTG   3309

TGAACCCTGCCAGGAGCAGGGCCTGGACCAGCAGGCCATGAATAGACATACTTGGTGAAGTGAACGGAGACTGAGGATG   3388

GCTCTGCTTCCACCGAGGGAGACACTAGTTGGCAAAGTGTCTAACCTCCCTTTTCCAGCCCATTGCTCAAGTCCCCCAG   3467

GCTGTGGACATGAGCTGGTGGGCAGAATGTTGTTGTTGAAGTCTGATTTTAGATTGATTTTTTAAAAAAAAAAAAAAAA   3546

AAAAAAAAAAAGGGCGGCCGC                                                             3567
```

FIGURE 33C

```
                   10        20        30        40        50        60
inputs GTC-GACCCACGCGTCCGCTCGAAGCGGGGACCCTCGCCCCGTCCTCGGCTGTCCAGTCCTCCTCCTCGC
       ::: ::::::::::::::     :::       :.:.::: :.    :: ::    .:
       GTCCGACCCACGCGTCCG-----AGC-------CACACCCTGAAGGTGGTTGGAAGG-------------
                 10        20                 30        40

     70        80        90       100       110       120       130
inputs AGACCCCGGCGGTTCCTACCCCAGGCCGCAGGGGAGACGGTGCCCCAAGGCAGGCTTCATA--TCCTGAA
       ::.     :. ::.:: ..  ::.:. :.:.::        ..: :::::..::: :.::. .  ::: ..:
       AGG----GAAGGATCTAGGTCCTGAGCACTGG------AATTCCCCAGAACAG-CATCTGGCTTCCCAGA
              50        60        70             80        90       100

       140       150       160       170       180       190       200
inputs CGCTGG-GATCCCCCA-GGACATTCCCTGGCCCCCAGGCCCCAGGTCCCAGGCCCCAGGGCTGAGCTGTG
       : :. :   . :: :::  :: .:  :::        ::.:     :.:     :.:::  ::: ::::. ::::
       CCCATGCTGGCCACCACTGATGTGTCCTT----CCGG----CTG----CTGGCTGCAGTGCTGTTCTGTT
          110       120       130                140       150       160

        210       220       230       240       250       260       270
inputs GGCAGGCCCCACCTGGCCTCTGCAATGTCACCGCCTCTGTGTCCCCTCCTTCTCCTGGCTGTGGGCCTGC
       :  .:: ::   ::::    ::..    ::   :::.:: ::::::::::: ::::::::  :.:::::::
       GTTGGGTGCCCTGTGGCA--GGCTTGTGCAATGCCACTCTGTCCCCTCCTCCTCCTGGCCCTAGGCCTGC
          170       180       190       200       210       220       230

        280       290       300       310       320       330       340
inputs GGCTGGCTGGAACTCTCAACCCCAGTGATCCCAATACCTGCAGCTTCTGGGAAAGCTTCACTACCACCAC
       : :::::::::::.:::::: :::.::::::::::. ::: : :::::::::::::::: :: :::::
       GTCTGGCTGGAACACTCAACTCCAATGATCCCAATGTCTGTACCTTCTGGGAAAGCTTCACCACGACCAC
          240       250       260       270       280       290       300

        350       360       370       380       390       400       410
inputs CAAGGAGTCCCACTCCCGCCCCTTCAGCCTGCTCCCCTCAGAGCCCTGCGAGCGGCCCTGGGAGGGCCCC
        :::::::::::    ::::::::::::::: :::   : ::: :::::::  ::::::::::.:.::::
       TAAGGAGTCCCACCTTCGCCCCTTCAGCCTGCCCCCAGCCGAGTCCTGCGACAGGCCCTGGGAAGACCCC
          310       320       330       340       350       360       370

        420        430            440         450        460       470
inputs CATACTTGC-CCCAGCCCACAAA---CT--CAGA---GGAAACTCCTGGCT-TCTAGGGATTCATTCTGC
       :: :: ::: : ::::: ::...    ::   :.::     :.: :  :.::   : .::. . .:.  :  :
       CACACCTGCGCTCAGCCTACGGTTGTCTACCGGACTGTGTACCGTCAGGTGGTGAAGATGGACTCCCGCC
          380       390       400       410       420       430       440

        480       490        500       510        520       530       540
inputs ATGGTCTGTGTCGGGGCTG-GAGTGCAGTGGCGAGATC-GTAGTGCACTGCAACCTCAAACAGGGAATGC
       . : :::   :.: :::: :.: : ..  .::::: : ::.:.:: ::: ..:: :: .:.:      :::
       CACGCCTG---CAGTGCTGTGGGGGTTACTACGAGAGCAGTGGAGC-CTGTGTCC-CACTCTG----TGC
             450       460       470       480       490       500

          550       560        570       580       590       600       610
inputs GCTTTCTATGCGCCCTCAGCCCAGAGTGTTGAGTGGTGCCCCTTCCCTG-GCCTCCCCTGGCCACACTGT
        :.   ..:: : ::.:.: :         : ::.:::  :: :::.  : : :::.       .::: .:::
       CCAGG-AGTGTGTCCACGGTC------GCTGTGTG--GCTCCTAATCGGTGCCAGTGTGCACCAGGCTGG
          510       520               530       540       550       560
```

Figure 34A

```
           620       630       640       650       660       670       680
inputs GGTGGTGAAGACGGACCACCGCCAGCGCCTGCAGTGCTGCCATGGCTTCTATGAGAGCAGGGGGTTCTGT
        : ::::: ::: :.       :::: :  .: .::::: :: ::: .  :.::.:..: . .:    :::
       CGGGGTGACGACTGT-----TCCAGTG--AG-TGTGCT-CC-TGGAA--TGTGGGGACCACAG----TGT
           570            580          590          600       610

           690       700       710       720       730         740       750
inputs GTCCCGCTCTGTGCCCAGGAGTGTGTCCATGGCCGTTGTGTGGCACCCA--ATCAGTGCCAATGTGTGCC
       :.:  ::::::   ::.    :::: : : .:: :::         : .::::  :   .: :   ..: : ::::
       GACAGGCTCTG---CCTC---TGTGGCAACAGCAGTTCCTGTGATCCCAGGAGTGGGGTGTGTTTTTGCC
        620        630          640       650       660        670       680

              760       770       780       790       800          810
inputs AGGCTGGCGGGGCGACGACTGTTCCAGTGCCCCGAACTGCCTTCAGCCCTGTACCCC--TGGCTACTATG
          :::::          : .:.:  ::    :::::: ::::::::::: ::   ::::  :::: ::::::
       CCTCTGGC-------CTGCAG--CC----CCCCGA-CTGCCTTCAGCCTTG--CCCCGATGGCCACTATG
                      690              700        710          720       730

     820       830       840       850       860       870       880
inputs GCCCTGCCTGCCAGTTCCGCTGCCAGTGCCATGGGGCACCCTGCGATCCCCAGACTGGAGCCTGCTTCTG
       : ::::::::::::::  . ::::: ::: :::::::: :::: :: ::::.:. ::::::::::::::
       GTCCTGCCTGCCAGTTTGATTGCCATTGCTATGGGGCATCCTGTGACCCCCGGGATGGAGCCTGCTTCTG
          740       750       760       770       780       790       800

     890       900       910       920       930       940       950
inputs CCCCGCAGAGAGAACTGGGCCCAGCTGTGACGTGTCCTGTTCCCAGGGCACTTCTGGCTTCTTCTGCCCC
       :::: :::.::::::.::.:::::                                  ::::::  ::::::::::::::::
       CCCCCCAGGGAGAACAGGACCCAG----------------------GGCACTGATGGCTTCTTCTGCCCC
          810       820                                830       840       850

     960       970       980       990      1000      1010      1020
inputs AGCACCCATCCTTGCCAAAATGGAGGTGTCTTCCAAACCCCACAGGGCTCCTGCAGCTGCCCCCCTGGCT
       :: ::  ::::::::::::::::::::      ::.. : :.::.::::::::::::::::: :: ::::
       AGAACTTATCCTTGCCAAAATGGAGGTGTTCCTCAGGGCTCTCAAGGCTCCTGCAGCTGCCCACCGGGCT
             860       870       880       890       900       910       920

    1030      1040      1050      1060      1070      1080      1090
inputs GGATGGGCACCATCTGCTCCCTGCCCTGCCCAGAGGGCTTTCACGGACCCAACTGCTCCCAGGAATGTCG
       ::::::: . :::::: :::::::: ::::::::::: :: :::::::::::::: .: :::::::::::
       GGATGGGTGTCATCTGTTCCCTGCCATGCCCAGAGGGTTTCCACGGACCCAACTGTACTCAGGAATGTCG
             930       940       950       960       970       980       990

    1100      1110      1120      1130      1140      1150      1160
inputs CTGCCACAACGGCGGCCTCTGTGACCGATTCACTGGGCAGTGCCGCTGCGCTCCGGGTTACACTGGGGAT
        :::::::: :: ::::: :::::: :.:: :::::::::::::.::: ::::: :: :: :  ::::::
       TTGCCACAATGGTGGCCTTTGTGACAGGTTTACTGGGCAGTGCCACTGTGCTCCTGGCTATATCGGGGAT
            1000      1010      1020      1030      1040      1050      1060

    1170      1180      1190      1200      1210      1220      1230
inputs CGGTGCCGGGAGGAGTGCCCGGTGGGCCGCTTTGGGCAGGACTGTGCTGAGACGTGCGACTGCGCCCCGG
       :::::::: ::.::::::::: ::::::::::: :: ::.:::::::::::::: :: ::::: :: :: :
       CGGTGCCGTGAAGAGTGCCCTGTGGGCCGCTTCGGTCAAGACTGTGCTGAGACCTGTGACTGTGCTCCTG
            1070      1080      1090      1100      1110      1120      1130

    1240      1250      1260      1270      1280      1290      1300
inputs ACGCCCGTTGCTTCCCGGCCAACGGCGCATGTCTGTGCGAACACGGCTTCACTGGGGACCGCTGCACGGA
       .::: :::::::: :: ::::: :::::.:::::::::::::::: ::::::::.:: ::::::::::: ::
       GCGCTCGTTGCTTTCCTGCCAATGGCGCGTGTCTGTGCGAACATGGCTTCACAGGCGACCGCTGCACTGA
            1140      1150      1160      1170      1180      1190      1200
```

Figure 34B

```
     1310      1320      1330      1340      1350      1360      1370
inputs TCGCCTCTGCCCCGACGGCTTCTACGGTCTCAGCTGCCAGGCCCCCTGCACCTGCGACCGGGAGCACAGC
        :: ::::: :: :: :::  ::: ::::: :::::::::.:  ::::::::::::::: .::.:::::
       GCGACTCTGTCCAGATGGCCGCTATGGTCTGAGCTGCCAAGATCCCTGCACCTGCGACCCAGAACACAGT
           1210      1220      1230      1240      1250      1260      1270

     1380      1390      1400      1410      1420      1430      1440
inputs CTCAGCTGCCACCCGATGAACGGGGAGTGCTCCTGCCTGCCGGGCTGGGCGGCCTCCACTGCAACGAGA
       ::::::::::::::.::: :::: :::::::::::::.:::.:: :::::::::::::::::::::::::
       CTCAGCTGCCACCCAATGCACGGCGAGTGCTCCTGCCAGCCAGGTTGGGCGGCCTCCACTGCAACGAGA
           1280      1290      1300      1310      1320      1330      1340

     1450      1460      1470      1480      1490      1500      1510
inputs GCTGCCCGCAGGACACGCATGGGCCAGGGTGCCAGGAGCACTGTCTCTGCCTGCACGGTGGCGTCTGCCA
       ::::::: ::::::::::: ::. : :: :::::::::::::: ::::: :::::::: :: :: ::::.
       GCTGCCCTCAGGACACGCACGGAGCCGGTTGCCAGGAGCACTGCCTCTGTCTGCACGGCGGTGTTTGCCT
           1350      1360      1370      1380      1390      1400      1410

     1520      1530      1540      1550      1560      1570      1580
inputs GGCTACCAGCGGCCTCTGTCAGTGCGCGCCGGGTTACACGGGCCCTCACTGTGCTAGTCTTTGTCCTCCT
        :: . :::::::::::: :.::: ::.:: :: :::::::: :::::::: ::::.:::::::::.:::
       CGCCGACAGCGGCCTCTGCCGGTGTGCACCTGGCTACACGGGACCTCACTGCGCTAATCTTTGTCCACCT
           1420      1430      1440      1450      1460      1470      1480

     1590      1600      1610      1620      1630      1640      1650
inputs GACACCTACGGTGTCAACTGTTCTGCACGCTGCTCATGTGAAAATGCCATCGCCTGCTCACCCATCGACG
       .:::: :: :: .::::::::::  : :.:::::: :::::::::::::: ::::::::.:: .::::::
       AACACTTATGGGATCAACTGTTCCTCCCACTGCTCCTGTGAAAATGCCATTGCCTGCTCTCCTGTCGACG
           1490      1500      1510      1520      1530      1540      1550

     1660      1670      1680      1690      1700      1710      1720
inputs GCGAGTGCGTCTGCAAGGAAGGTTGGCAGCGTGGTAACTGCTCTGTGCCCTGCCCACCCGGAACCTGGGG
       ::. ::::.::::::::::::::::::::::::::::::::::::::::: :: :: :: ::::::::
       GCACGTGCATCTGCAAGGAAGGTTGGCAGCGTGGTAACTGCTCTGTGCCCTGTCCCCCTGGCACCTGGGG
           1560      1570      1580      1590      1600      1610      1620

     1730      1740      1750      1760      1770      1780      1790
inputs CTTCAGTTGCAATGCCAGCTGCCAGTGTGCCCATGAGGCAGTCTGCAGCCCCCAAACTGGAGCCTGTACC
       :::::::::::::::::: :::::::::::::: :::: :::::::::::::::::::::::::::::
       CTTCAGTTGCAATGCCAGTTGCCAGTGTGCCCACGAGGGAGTCTGCAGCCCCCAAACTGGAGCCTGTACT
           1630      1640      1650      1660      1670      1680      1690

     1800      1810      1820      1830      1840      1850      1860
inputs TGCACCCCTGGGTGGCATGGGGCCCACTGCCAGCTGCCCTGTCCGAAGGGGCAGTTTGGAGAAGGTTGTG
       ::::::::::::::::.::::::  :::::::::.:: :: :: ::::::::.::::::::.::::::::::
       TGCACCCCTGGGTGGCGTGGGGTTCACTGCCAACTTCCGTGCCCGAAGGGACAGTTTGGTGAAGGTTGTG
           1700      1710      1720      1730      1740      1750      1760

     1870      1880      1890      1900      1910      1920      1930
inputs CCAGTCGCTGTGACTGTGACCACTCTGATGGCTGTGACCCTGTTCATGGACGCTGTCAGTGCCAGGCTGG
       :::::  :::::::::::::::::: ::::::::::::::::::::::::.::: :..:: ::::::::
       CCAGTGTCTGTGACTGTGACCACTCCGATGGCTGTGACCCTGTTCATGGACACTGCCGATGTCAGGCTGG
           1770      1780      1790      1800      1810      1820      1830

     1940      1950      1960      1970      1980      1990      2000
inputs CTGGATGGGTGCCCGCTGCCACCTGTCCTGCCCTGAGGGCTTATGGGGAGTCAACTGTAGCAACACCTGC
       :::::::::: .: :: :::::::::: : :::::.:::::::::.:::::::: :::::: ::::: .::::
       CTGGATGGGCACACGTTGCCACCTGCCTTGCCCAGAGGGCTTTTGGGGAGCCAACTGCAGCAATGCCTGT
           1840      1850      1860      1870      1880      1890      1900
```

Figure 34C

```
           2010      2020      2030      2040      2050      2060      2070
inputs ACCTGCAAGAATGGGGGCACCTGTCTCCCTGAGAATGGCAACTGCGTGTGTGCACCCGGATTCCGGGGCC
       :::::::::::::: ::::: ::: : :::::::: :::::::: ::::: ::::: ::.::: :.::::
       ACCTGCAAGAATGGTGGCACTTGTGTACCTGAGAACGGCAACTGTGTGTGCGCACCAGGGTTCAGAGGCC
           1910      1920      1930      1940      1950      1960      1970

           2080      2090      2100      2110      2120      2130      2140
inputs CCTCCTGCCAGAGATCCTGTCAGCCTGGCCGCTATGGCAAACGCTGTGTGCCCTGCAAGTGCGCTAACCA
       ::::::::::::::. :::: : :::::: ::::::::::::::::::::::::::::::::::. :::::
       CCTCCTGCCAGAGGCCCTGCCCGCCTGGTCGCTATGGCAAACGCTGTGTGCCCTGCAAGTGCAACAACCA
           1980      1990      2000      2010      2020      2030      2040

           2150      2160      2170      2180      2190      2200      2210
inputs CTCCTTCTGCCACCCCTCGAACGGGACCTGCTACTGCCTGGCTGGCTGGACAGGCCCCGACTGCTCCCAG
        :: : ::::::::: :::.: :::::::::: :::::::::.::::::::::::::: :::::::  :.
       TTCTTCCTGCCACCCGTCGGATGGGACCTGCTCCTGCCTGGCAGGCTGGACAGGCCCTGACTGCTCTGAA
           2050      2060      2070      2080      2090      2100      2110

           2220      2230      2240      2250      2260      2270      2280
inputs CCATGCCCTCCAGGACACTGGGGAGAAAACTGTGCCCAGACCTGCCAATGTCACCATGGTGGACCTGCC
        :::: :: ::::: ::::::::: . :: ::  ::::. ::::::::.::::: :::::::  :::::::
       TCATGTCCCCCAGGCCACTGGGGACTCAAATGCTCCCAACCCTGCCAGTGTCATCATGGTGCCACCTGCC
           2120      2130      2140      2150      2160      2170      2180

           2290      2300      2310      2320      2330      2340      2350
inputs ATCCCCAGGATGGGAGCTGTATCTGCCCCCTAGGCTGGACTGGACACCACTGCTTAGAAGGCTGCCCTCT
       : ::::::::::::::::::::.:::::  :: :::::::::::::: : :::::: .::::::::::::.
       ACCCCCAGGATGGGAGCTGTGTCTGCATCCCAGGCTGGACTGGACCCAACTGCTCGGAAGGCTGCCCATC
           2190      2200      2210      2220      2230      2240      2250

           2360      2370      2380      2390      2400      2410      2420
inputs GGGGACATTTGGTGCTAACTGCTCCCAGCCATGCCAGTGTGGTCCTGGAGAAAAGTGCCACCCAGAGACT
       ..:.: .:::::::  ::::::::::::: ::: :::::::.:::::::::.:.:::::::::::::::::
       AAGAATGTTTGGTGTCAACTGCTCCCAGCTATGTCAGTGTGATCCTGGAGAGATGTGCCACCCAGAGACT
           2260      2270      2280      2290      2300      2310      2320

           2430      2440      2450      2460      2470      2480      2490
inputs GGGGCCTGTGTATGTCCCCCAGGGCACAGTGGTGCACCTTGCAGGATTGGAATCCAGGAGCCCTTTACTG
       ::::: :: :: :::::::::::.:::::::::::.:  ::::...: :: : ::::::: :::: :: .
       GGGGCTTGCGTCTGTCCCCCAGGACACAGTGGTGCGCACTGCAAAGTGGCAGCCAGGAGTCCTTCACCA
           2330      2340      2350      2360      2370      2380      2390

           2500      2510      2520      2530      2540      2550      2560
inputs TGATGCCGACCACTCCAGTAGCGTATAACTCGCTGGGTGCAGTGATTGGCATTGCAGTGCTGGGGTCCCT
       :.::::: :::.::::.::..    ::::::::.::::::::: ::::::::::::::::::::::::.::::
       TAATGCCCACCTCTCCTGTGATCCATAACTCACTGGGTGCCGTGATTGGCATTGCAGTGCTGGGGACCCT
           2400      2410      2420      2430      2440      2450      2460

           2570 -    2580      2590      2600      2610      2620      2630
inputs TGTGGTAGCCCTGGTGGCACTGTTCATTGGCTATCGGCACTGGCAAAAAGGCAAGGAGCACCACCACCTG
       ::::::.:::::::::.:::::::: ::::::::: ::.::::::::::::.:::::::::.::  : ::: ::
       TGTGGTGGCCCTGGTAGCACTGTTTATTGGCTACCGACACTGGCAAAAGGGCAAGGAACATGAGCACTTG
           2470      2480      2490      2500      2510      2520      2530

           2640      2650      2660      2670      2680      2690      2700
inputs GCTGTGGCTTACAGCAGCGGGCGCCTGGACGGCTCCGAGTATGTCATGCCAGATGTCCCTCCGAGCTACA
       ::.:::::::::::::  ::::: ::::: :::::::: :: :::::::::::::::: ::::::::::::
       GCAGTGGCTTACAGCACTGGGCGACTGGATGGCTCCGATTACGTCATGCCAGATGTCTCTCCGAGCTACA
           2540      2550      2560      2570      2580      2590      2600
```

Figure 34D

```
          2710      2720      2730      2740      2750      2760      2770
inputs GTCACTACTACTCCAACCCCAGCTACCACACCCTGTCGCAGTGCTCCCCAAACCCCCCACCCCCTAACAA
       :::::::::: :::::::: :::::::::: ::::: ::::: :: ::.::::: ::::::::::::::
       GTCACTACTATTCCAACCCTAGCTACCACACACTGTCTCAGTGTTCTCCTAACCCTCCACCCCCTAACAA
           2610      2620      2630      2640      2650      2660      2670

          2780         2790      2800      2810      2820      2830      2840
inputs GGTTCCAGGC---CCGCTCTTTGCCAGCCTGCAGAACCCTGAGCGGCCAGGTGGGCCCAAGGGCATGAT
       :.::::::::   : ::: :::: ::::    :::.   :::::::::::.. .:.. :::.:::::..:::
       GATTCCAGGCAGTCAGCTGTTTGTCAGCTCCCAGGCATCTGAGCGGCCAAACAGAAACCATGGGCGAGAT
           2680      2690      2700      2710      2720      2730      2740

            2850      2860      2870      2880      2890      2900      2910
inputs AACCACACCACCCTGCCTGCTGACTGGAAGCACCGCCGGGAGCCCCCT-CCAGGGCCTCTGGACAGGGGG
       ::::::.:::: ::::: :::::::::::::::: :::::: ::: :  :::.::   :   :. .::
       AACCACGCCACACTGCCCGCTGACTGGAAGCACCGACGGGAGTCCCATGACAGAGC---TTTCCTCAGGC
           2750      2760      2770      2780      2790         2800

             2920      2930                2940      2950      2960      2970
inputs AGCAGCCGCCTGGACCGAAG------CTACAGCTATAGCTACAGCAATGGCCCAGGCCCATTCTACGATA
       : :::::.:::::::::::::      ::. ::::::.:: :::: :::::::::.:: :::::::.  :::
       ACCAGCCACCTGGACCGAAGGTATAGCTGTAGCTATGGCCACAGGAATGGCCCGGGGCCATTCTGTCATA
    2810      2820      2830      2840      2850      2860      2870

         2980      2990      3000      3010      3020      3030      3040
inputs AAGGGCTCATCTCTGAAGAGGAGCTCGGGGCCAGTGTGGCTTCCCTGAGCAGTGAGAACCCATATGCCAC
       :::: : ::::::::::::.:..:: ::::: :: :: .  :::::::::::::::::::: ::::: ::
       AAGGTCCCATCTCTGAAGAAGGACTAGGGGCAAGCGTTATGTCCCTGAGCAGTGAGAACCCCTATGCGAC
    2880      2890      2900      2910      2920      2930      2940

         3050      3060      3070      3080      3090      3100      3110
inputs CATCCGGGACCTGCCCAGCTTGCCAGGGGGCCCCCGGGAGAGCAGCTACATGGAGATGAAAGGCCCTCCC
       ::::::.::::::::::.:: ::::.::::. :::::.::.::::::::: .::::::::::::::::::::
       CATCCGAGACCTGCCCGGCCTGCCTGGGGAACCCCGAGAAAGCAGCTATGTGGAGATGAAAGGCCCTCCA
    2950      2960      3070      2980      2990      3000      3010

         3120      3130      3140      3150      3160      3170      3180
inputs TCAGGATCTGCCCCCAGGCAGCCTCCTCAGTTTTGGGACAGCCAGAGGCGGCGGCAACCCCAGCCACAGA
       :::: .::: :::::::::::::::: :::   :   ::::::: :::    :.::.:::::   ::: :.::::
       TCAGTGTCTCCCCCCAGGCAGCCTCTTCATCTCCGGGACAGGCAG---CAGCAGCAACTGCAGTCTCAGA
    3020      3030      3040      3050      3060          3070      3080

         3190      3200      3210      3220      3230      3240      3250
inputs GAGACAGTGGCACCTACGAGCAGCCCAGCCCCCTGATCCATGACCGAGACTCTGTGGGCTCCCAGCCCCC
       ::::::: :::::::: ::::::::::   ::: ::: ::.:.:  .::: :::::::::::: .::::::
       GAGACAGCGGCACCTATGAGCAGCCCACTCCCTTGAGCCGTAATGAAGAGTCTGTGGGCTCCATGCCCCC
       3090      3100      3110      3120      3130      3140      3150

         3260      3270      3280      3290      3300      3310      3320
inputs TCTGCCTCCGGGCCTACCCCCCGGCCACTATGACTCACCCAAGAACAGCCACATCCCTGGACATTATGAC
       ::: :::::::::::.:: ::::::::::::::::::.:::::.:::::::::::::::::::: ::::::
       TCTTCCTCCGGGCCTGCCACCCGGCCACTATGACTCGCCCAAAAACAGCCACATCCCTGGACACTATGAC
       3160      3170      3180      3190      3200      3210      3220

         3330      3340      3350      3360      3370      3380      3390
inputs TTGCCTCCAGTACGGCATCCCCCATCACCTCCACTTCGACGCCAGGACCGTTGAGGAGCCAGGATGGTAT
       :::::::::::::::::::: ::::::::::::    ::.:::::::::::: :::::::::: :::::::
       TTGCCTCCAGTACGGCATCCTCCATCACCTCCATCCCGGCGCCAGGACCGCTGAGGAGCCAGCATGGTAT
       3230      3240      3250      3260      3270      3280      3290
```

Figure 34E

```
    3400      3410      3420      3430      3440      3450      3460
inputs GGCAGAGGCCAGCACACCTGGCTGTTGCTGCTCAAGGCTGGGGACAGAGCCTAGTGTACCCCTGCCAGGA
       ::  :::                                      ::.:::: :::.:::: ::::::::
       GG--GAG--------------------------------------AGTGCCT-GTGAACCC-TGCCAGGA
          3300                                                3310      3320

    3470      3480      3490      3500      3510      3520      3530
inputs GCAGGGAGTGGACCGGCAGGCTGTGAACATGAACAACGCTTAACAGAGCAAGTGATGGGAGCCTTGTTCC
       ::::::  :::::::.::::::      ::::::      :..::: :
       GCAGGGCCTGGACCAGCAGGC------CATGAA------TAGACATA-----------------------
          3330      3340            3350

    3540      3550      3560      3570      3580      3590      3600
inputs TGGGTTCTACCATGGGAGACGCTGATCAGCAGGATGCCTGGCTCCCTTTCCCAACCCACTGCTCCCAAGG
                 :.:::       :::.
       ----------CTTGG-------TGAA--------------------------------------------
                3360

    3610      3620      3630      3640      3650      3660      3670
inputs CCTCCAGGGCCCTGTGTACATAAACTGGTGGGTTGGAAGTTGCTGGGTAACTCTGATTTCAGACATGCGT
                    :::.::. :.::::  .::.:::
       -------------GTGAACGGAGACTG--AGGATGG----------------------------------
                  3370      3380

    3680      3690      3700      3710      3720      3730      3740
inputs GTGGGGTACCTTTTCTGTGCATGCTCAGCCTGGGCTCTGTGCGTGTGTGTTTCTGTGATTTTAGAAGG
                              :::.::
       -----------------------CTCTGC-----------------------------------------
                           3390

    3750      3760      3770      3780      3790      3800      3810
inputs GTACCAGGCAGGTTCTGTCCTAGGGCACTTACCATTTAGTAGGGAGATGGAACCAACCCAATTAACTCTA
        :.:::             :: ::::                                    ..::.:::
       -TTCCA-------------CCGAGGG------------------------------------AGACACTA
          3400                                                          3410

    3820      3830      3840      3850      3860      3870      3880
inputs GCAATAGCCTCCTAACTGGCCTCCTCCATTGATTCAGTGAACCTTCCAATGCATGGCTCATAATTTCAAA
       :                                                   .::::
       G---------------------------------------------------TTGGC-------------
                                                           3420

    3890      3900      3910      3920      3930      3940      3950
inputs ATACAGGCTGGTTAGTTACTCCCTACCTGAAAGCCTTCATAGGTGCCTCTTTGCTCTTCTGCCAGTATCA
                                    ::::
       -----------------------------AAAG-------------------------------------

    3960      3970      3980      3990      4000      4010      4020
inputs AAACTTTTGAAGGCCTTAAAGGCCCTGCTTTGCCTGGCCCATCTGTCTCTCCAGCCTCACCTTGAACTGT
                                                 :::::
       ------------------------------------------TGTCT-----------------------
                                                  3430

    4030      4040      4050      4060      4070      4080      4090
inputs GTTCCTGTCACTGCACGCCAGTCACACCGGCCTCTAGGTCCTCCTGTAGGCCACTCTTCTTTCTGGCACA
                                            ..:::::
       -------------------------------------AACCTCC--------------------------
```

Figure 34F

```
   4100      4110      4120      4130      4140      4150      4160
inputs GGGACCTGCACACCTGGAGTGCCCTTCCTCCCCCACTCGCCTGTTCACCCCTGCTTTTCCTTTACACCTC
                                                            :::::::
       -----------------------------------------------------CTTTTCC----------
                                                         3440

   4170      4180      4190      4200      4210      4220      4230
inputs CTCCTCAGGGAAGTGCCCACCCTCCGTACATCTTTCACAGCCCTGATTGCAGCTGTGTTCACTCACCAGG
                                             :::::  :::::.               ::.:
       --------------------------------------AGCCC--ATTGCT---------------CAAG
                                                 3450

   4240      4250      4260      4270      4280      4290      4300
inputs TACCTGCAGAAGGCCTACAGGGTGCCAGGCACTTCTTTAATGGGTTCTTTCTTTATGTGATTATTTGATT
       :
       T---------------------------------------------------------------------
   3460

   4310      4320      4330      4340      4350      4360      4370
inputs AATCTCTGCCTCCCCCACTAGACTGTAAGCTCCCTGAAGGCAAGAATCCTGTGCTTATGCTCAATATTAG
                  ::::::
       -----------CCCCCA-----------------------------------------------------

   4380      4390      4400      4410      4420      4430      4440
inputs CTCTCCCTTGGCACAGAGTAGGCACTCAACAAATGCTCCCCAAAAGGCTGAGTGGCTGACTGAATTAAGT
                                                    :::::.:
       ---------------------------------------------GGCTGTG------------------
                                                    3470

   4450      4460      4470      4480      4490      4500      4510
inputs ACCAGTGACATGCAGTAACTGCTAAGATAGATGAGCCATCTGTATGCTCTGACAGTTACAGACTGAATAA
             ::::::
       ------GACATG----------------------------------------------------------

   4520      4530      4540      4550      4560      4570      4580
inputs GTTGGAGACTTCCCTAAAGGGTGGCATTTCCCCAGGGTAACAACGCAGAGCTCAGGTGTGGGAAGGTGCC

       ----------------------------------------------------------------------

   4590      4600      4610      4620      4630      4640      4650
inputs AGGGGCAGGGGTGCAGAGGGGCTGAGGCTGAGGGGGGTGCAGAGGCTGGAGAAAGGATAACAGGAGAGAG
                                                  .:::::.:.
       -------------------------------------------AGCTGGTGG------------------
                                             3480

   4660      4670      4680      4690      4700      4710      4720
inputs TATACAGGCATGCCTTGATTTATTGCACTTCACAGGTAGCAGAATTTTTAAAGAAATTGAAGGTTTTGGG
                                              ::::::: ::    :...::::::
       --------------------------------------GCAGAATGTT----GTTGTTGAAG--------
                                           3490          3500

   4730      4740      4750      4760      4770      4780      4790
inputs ACATATATGTGACAGCAATAGGTTAAGAAAAGCAAAGCAGAGAAATTGAAGATTTGTGTCAACACTGCTT

       ----------------------------------------------------------------------
```

Figure 34G

```
     4800      4810      4820      4830      4840      4850      4860
inputs TAAGCAAATCTGTTGGCACCATTTTTCCAATAGCATGTGCCCATTTTGGGTCTCTACATTGCATTTTGGT
               ::::                              :::::.:.:
       --------TCTG------------------------------ATTTTAGAT-------------------
             3510                                    3520

     4870      4880      4890      4900      4910      4920      4930
inputs AATTGCTTGCAATATTTCAAGCATTTTCATTGTTATTATATGTGTTATAGTGATCTGTGATCAGTGATCT

       ----------------------------------------------------------------------

     4940      4950      4960      4970      4980      4990      5000
inputs TTGATATATTATTGTAATTGTTTCGGGGCGCCATGAACCGCACCCATATAACACGGTAAACTTAATCAGC
        ::::.:.:::....::.
       -TGATTTTTTAAAAAAAA----------------------------------------------------
             3530

     5010      5020      5030
inputs AAAAAAAAAAAAAAAAAAAAGGGCGGCCG-
       :::::::::::::::::::::::::::::
       AAAAAAAAAAAAAAAAAAAAGGGCGGCCGC
     3540      3550      3560
```

Figure 34H

```
                 10                  20             30                     40             50
inputs GTC-GACCCACGCGTCCGG---TGACCCTGTTCATGGACAGT-----GCCGATGTCAGG---CTGGT---
       ::: ::::::::::::::.    . :::::::.. .::::. .:.      :   .. .::..:    :::.
       GTCCGACCCACGCGTCCGAGCCACACCCTGAAGGTGGTTGGAAGGAGGGAAGGATCTAGGTCCTGAGCAC
               10             20            30            40            50           60          70

          60              70                 80              90                100               110
inputs TGGATGGGCACA-CGCTGCCAC---CTGCCTTG-CCCGGA--GG--GCTTTTGGGGAG-CCAAC-TGCAG
       ::::.    : ::   .:.:: .:     :: :: .: :::. .    ::    .: . ::. :.: ::..:  ::.:
       TGGAATTCCCCAGAACAGCATCTGGCTTCCCAGACCCATGCTGGCCACCACTGATGTGTCCTTCCGGCTG
               80             90            100           110           120          130          140

            120             130            140              150              160             170
inputs -TAACACCTGTACC-TGCAAGAATGGTGGTACCTGTG--TGTCT-GAGAATGGCAACTGCGTGTGCGCAC
       :..:. :.::.:  : :..  ..: : :     :::::: .: :: :.: :. :: ::: : ::: : :.:
       CTGGCTGCAGTGCTGTTCTGTTGTTGGGTGCCCTGTGGCAGGCTTGTGCAATGCCACT-C-TGTCCCCTC
                150           160            170            180            190           200

     180             190               200            210               220             230
inputs CAG----GGTTCCGAGGCCC-CTCCTGCCAGAGGCCCTGCCCGCC--TGGTCGCTATGGCAA-AC--GCT
       :.         :  :: :::::  :   ::: :.:...: ::     : ::   ::.:: :.::: :.. ::    ::
       CTCCTCCTGGCCCTAGGCCTGCGTCTGGCTGGAACACTCAACTCCAATGATCCCAATGTCTGTACCTTCT
       210           220            230            240           250           260          270

     240             250                              260              270              280
inputs GTGT--GCAATGC-----------AAGTGT---AACAACAACCATTCTTCCTGCCACCCATCG-------
       : :.  ::.. .:                   :.:.::    : :..:. :: :::. :::::: :: . ::
       GGGAAAGCTTCACCACGACCACTAAGGAGTCCCACCTTCGCCCCTTCAGCCTGCCCCCAGCCGAGTCCTG
       280           290            300            310           320           330          340

         290                       300                310              320              330
inputs -GACGGGACCTG-------CTCCT-GCCTG---GCGGGCTG-GACAGGC--CCTGACTGC--TCCG--AG
        :::.:: ::::         : ::. .::::     :.: ::. :. .: :    :: :::::     .:::   ::
       CGACAGGCCCTGGGAAGACCCCCACACCTGCGCTCAGCCTACGGTTGTCTACCGGACTGTGTACCGTCAG
       350           360            370            380           390           400          410

                340                 350                       360                  370
inputs GC-------ATG---TCCC--CCAGGCCA------CTGGGG-----ACT-CAAATGCT------CC----
       :         :::    ::::  ::: :::.        ::: ::       ::: :.:..::.          ::
       GTGGTGAAGATGGACTCCCGCCCACGCCTGCAGTGCTGTGGGGGTTACTACGAGAGCAGTGGAGCCTGTG
       420           430            440            450           460           470          480

            380                  390               400                         410
inputs --CAACTCTG---CCAG-----TGTCATCA-----TG-GTGGGACCT-------GCCA-----CCCC---
         : ::::::    ::::        :::: .:.      :: :::: .:::          ::::       : ::
       TCCCACTCTGTGCCCAGGAGTGTGTCCACGGTCGCTGTGTGGCTCCTAATCGGTGCCAGTGTGCACCAGG
       490           500            510            520           530           540          550
```

Figure 35A

```
           420         430          440      450              460
inputs CAGGATGGGAG----CTGTATC------TGCACGCCAGGCTGGACTGGACC-CAA----------CTGC
       :.::  ::::.:   ::::. :      :: .: ::.:: . :.  ::::: ::.          ::::
       CTGGCGGGGTGACGACTGTTCCAGTGAGTGTGCTCCTGGAATGTGGGACCACAGTGTGACAGGCTCTGC
      560       570       580       590       600       610       620

             470                 480       490                      500
inputs TTGGAAGGCTGC-------CCA------CCAAGAATGTTTGGTGT--------------CAACTGCTCC
        :  ..::::..:       ::.      :::.::.::    ::::              :..:.:: ::
       CTCTGTGGCAACAGCAGTTCCTGTGATCCCAGGAGTGGGGTGTGTTTTTGCCCCTCTGGCCTGCAGCCCC
      630       640       650       660       670       680       690

                    510             520              530
inputs C-----------AGCTATGTC--AGTG-------TGATCT-------CGGAGAGATG-----TGC-----
       :           ::: .:: :  ..::       ::.::       :.:. .:::     :::
       CCGACTGCCTTCAGCCTTGCCCCGATGGCCACTATGGTCCTGCCTGCCAGTTTGATTGCCATTGCTATGG
      700       710       720       730       740       750       760

          540                550       560       570                580
inputs --CACCCAGAGAC--------TGGGGCTTGTGTCTGTCCCCCAGG---------ACACAG-----TGGTG
         :: ::.:.:::        :::.:: ::  :::: ::::::::         :: :::     ::.::
       GGCATCCTGTGACCCCCGGGATGGAGCCTGCTTCTGCCCCCCAGGGAGAACAGGACCCAGGGCACTGATG
      770       780       790       800       810       820       830

                                 590       600          610       620
inputs -------CAGAC--------------TGCAAAATGGAAG----CC--AGGAGTC-CTT--CACCATAA-
              :.: :              ::: :::. :::.:    ::  :::. :: :..  :.::.  :
       GCTTCTTCTGCCCCAGAACTTATCCTTGCCAAAATGGAGGTGTTCCTCAGGGCTCTCAAGGCTCCTGCAG
      840       850       860       870       880       890       900

                                 630           640                    650
inputs -TGCCCACC-----------------TCT---CCCG---TGACCCATAA------CTC-------ACTGG
        ::::::::                 :::   :::    . .:::: :.      :.:       ::::
       CTGCCCACCGGGCTGGATGGGTGTCATCTGTTCCCTGCCATGCCCAGAGGGTTTCCACGGACCCAACTGT
      910       920       930       940       950       960       970

          660       670       680       690           700           710
inputs GTGCAGTGATTGGCATTGCAGTACTGGGAACCCTCGTG----GTGGCCCTGATAG---CACTGTTCAT-T
       .  ::: ::.:: :.:::: . : ::: ..:: : :::     ::    :   :. ::   :::::: :.  :
       ACTCAG-GAATGTCGTTGCCACAATGGTGGCCTTTGTGACAGGTTTACTGGGCAGTGCCACTGTGCTCCT
      980       990      1000      1010      1020      1030      1040

         720                               730                  740
inputs GGCTA----------------CCG----------CCAGTGG-----------CAAAA--GGGCAAGGAACA
       :::::                :::          ::.::::           :::.:  : ::...:: :.
       GGCTATATCGGGGATCGGTGCCGTGAAGAGTGCCCTGTGGCCGCTTCGGTCAAGACTGTGCTGAGACCT
      1050      1060      1070      1080      1090      1100      1110

           750       760                  770       780         790
inputs -----TGAGCACTTGGCA---GTGGCTTAC--------AGCACTGGGCGG--CTGG-ATGGCTCTGATTA
            ::.::.: ::::.    :: :::::.:        .::.:  : : :   : .. ::::::  . . .
       GTGACTGTGCTCCTGGCGCTCGTTGCTTTCCTGCCAATGGCGCGTGTCTGTGCGAACATGGCTTCACAGG
      1120      1130      1140      1150      1160      1170      1180
```

Figure 35B

```
        800        810        820         830         840      850
inputs CGTCA--TGC-CAGAT-GTCTCT--CCGA------GCTATAGTCACTACTACT------CCAACCCCAGC
       ::.:   ::: :.:: :.:::: :::..      ::::::.:::. ..:::.:      ::..: ::.::
       CGACCGCTGCACTGAGCGACTCTGTCCAGATGGCCGCTATGGTCTGAGCTGCCAAGATCCCTGCACCTGC
       1190      1200      1210      1220      1230      1240      1250

               860       870       880          890        900
inputs TACC--ACACACTGTCTCAGTGTTCTCCTAACCCCCCGC-----CCCCTAACA---AGGTCC--CAGGCA
        :::  . ::::.:::::::    :.:: ::  : : ::      : :::. ::   ::::    :.:::
       GACCCAGAACACAGTCTCAGCTGCCACCCAATGCACGGCGAGTGCTCCTGCCAGCCAGGTTGGGCGGGCC
       1260      1270      1280      1290      1300      1310      1320

         910       920       930          940          950
inputs G--TCAGCT-CTTTGTCAGCTCTCAGGCC-C---CTGAGC----GGCCA--AGCAGAGCC--------CA
         .:.::. : . . :.:: :::::: : :    : ::::     ::::  :::: .:::        ::
       TCCACTGCAACGAGAGCTGCCCTCAGGACACGCACGGAGCCGGTTGCCAGGAGCACTGCCTCTGTCTGCA
       1330      1340      1350      1360      1370      1380      1390

        960       970       980        990     1000         1010
inputs CGGGCGTGAGAACCATACCACACTGC--CCGCTGACTGGAAGCACC--GC----CGGGAGCCC------C
       :::  :::. ..::. .::. ::.::  :: ::: : : ..:::::  ::    ::::: : :      :
       CGGCGGTGTTTGCCTCGCCG-ACAGCGGCCTCTGCCGGTGTGCACCTGGCTACACGGGACCTCACTGCGC
       1400      1410       1420      1430      1440      1450      1460

          1020      1030                   1040      1050          1060
inputs ATGACAGAGGC-GCCAGCCAC------------CTGGACCGAA-GCTATAGCTGTA-----GCTATAGCC
       ....:. .: : .::.. :::            ::: .::     .::. . ::::.     :: ::.:::
       TAATCTTTGTCCACCTAACACTTATGGATCAACTGTTCCTCCCACTGCTCCTGTGAAAATGCCATTGCC
        1470      1480      1490      1500      1510      1520      1530

               1070      1080             1090      1100      1110
inputs A------CAGG-AATGGCCCAGG--AC--CATT-----CTGTCATAAAGGTCCCATCTCTGAA----GA-
       .       :.:  .: ::: :. :   .:  ::.       :: ::  ..:::  :. ::::::..    :.
       TGCTCTCCTGTCGACGGCACGTGCATCTGCAAGGAAGGTTGGCAGCGTGGTAACTGCTCTGTGCCCTGTC
        1540      1550      1560      1570      1580      1590      1600

              1120      1130      1140      1150              1160
inputs -----GGGACTAGGGGCAAGCGTTA-TGTCCCTGA-GCAGTGAGAACCC-CTA-------TGCTACC---
            :: :: .:::::..   :::. ..: ::.:. :: . :.:..::: : :         :::..::
       CCCCTGGCACCTGGGGCTTCAGTTGCAATGCCAGTTGCCA-GTGTGCCCACGAGGGAGTCTGCAGCCCCC
        1610      1620      1630      1640      1650      1660      1670

         1170                1180      1190              1200        1210
inputs -ATCCGAGACCTG---------CCCAGCCTGCC-TGGGGAAC----CC------CGAG---AAAGTGGCT
        :.: :...::::         ::::.: :: : ::::::..:     ::       ::.:   .::: :.:.
       AAACTGGAGCCTGTACTTGCACCCCTGGGTGGCGTGGGGTTCACTGCCAACTTCCGTGCCCGAAGGGACA
         1680      1690      1700      1710      1720      1730      1740

            1220      1230         1240      1250           1260
inputs ATGTGGAGATGAAAGGACC----TCCAT--CAGTGTCCCCTCCCA-GGCAGT----CTCTTCAT-----C
       .: :::.::.:...: .::    :: .:  :.:::.:: :::: : :::.::     :: :::::     :
       GTTTGGTGAAGGTTGTGCCAGTGTCTGTGACTGTGACCACTCCGATGGCTGTGACCCTGTTCATGGACAC
         1750      1760      1770      1780      1790      1800      1810
```

Figure 35C

```
           1270      1280                 1290      1300      1310
inputs T-CCGG-GACAGGCAG-CAG----------CGG---CAACTGC--AGCCACAGAGGG--ACAGCGGCACC
       : :::. :.::::::.: :.:          ::    :: ::::  .::: :::::::  . .: :: .::
       TGCCGATGTCAGGCTGGCTGGATGGGCACACGTTGCCACCTGCCTTGCC-CAGAGGGCTTTGGGGAGCC
        1820      1830      1840      1850      1860      1870      1880

            1320            1330                  1340      1350
inputs TA-TG-AGCA--GCC-----CAGC-----------CCCTTGAG--CCATAATGAAGAGTCTTTGGG----
       .: :: ::::  :::     :.::           : ::::.:  ::. :... .: ..:: :: :
       AACTGCAGCAATGCCTGTACCTGCAAGAATGGTGGCACTTGTGTACCTGAGAACGGCAACTGTGTGTGCG
         1890      1900      1910      1920      1930      1940      1950

   1360             1370      1380      1390      1400
inputs CTCCA-----C----GCCCCCGCTTCCTCCAGGCCTGCC-TCCTGGTCACTACGACT--C--------CC
       :.:::     :    :::::  :: ::.  .: ::::::  ::::::::.::: :.:.  :        ::
       CACCAGGGTTCAGAGGCCCCTCCTGCCAGAGGCCCTGCCCGCCTGGTCGCTATGGCAAACGCTGTGTGCC
         1960      1970      1980      1990      2000      2010      2020

   1410             1420      1430                  1440           1450
inputs C--CAAG---AACAGCCATA-TCCCTG------GAC-------ACTATGACTTGCCT--C---CAGTAC-
       :  ::::   ::::.::::. : ::::      :.:       :: .  .: :::::  :   :.: ::
       CTGCAAGTGCAACAACCATTCTTCCTGCCACCCGTCGGATGGGACCTGCTCCTGCCTGGCAGGCTGGACA
         2030      2040      2050      2060      2070      2080      2090

           1460                1470                        1480
inputs GGC---ATC--CTC-----CAT---CCCCT--CCA----------------TCCCGGC---GCCAG-GAC
       :::    ..:  :::     :::   ::::.  :::                ::::..:   ::::: :.:
       GGCCCTGACTGCTCTGAATCATGTCCCCCAGGCCACTGGGGACTCAAATGCTCCCAACCCTGCCAGTGTC
         2100      2110      2120      2130      2140      2150      2160

   1490       1500            1510      1520      1530                1540
inputs CGC-TGAAGA-GCCGGCAT-----GGTATGGGAGC-GTGCCTATGTACCTTGC----CAGGA-------G
        : ::..:  .:: :: .      .: :::::::: ::: ::. .: ::. ::    :.:::       :
       ATCATGGTGCCACCTGCCACCCCCAGGATGGGAGCTGTGTCTGCATCCCAGGCTGGACTGGACCCAACTG
         2170      2180      2190      2200      2210      2220      2230

             1550                   1560           1570                1580
inputs CAGGGACTG--GACCAGCAGG------------CCACG------AACAGAAACA-------CTTGGTGAA
       :. :::  :  : ::: ::.:             : ::      :.:.....::       : :::.::.
       CTCGGAAGGCTGCCCATCAAGAATGTTTGGTGTCAACTGCTCCCAGCTATGTCAGTGTGATCCTGGAGAG
         2240      2250      2260      2270      2280      2290      2300

             1590      1600      1610          1620      1630
inputs GTGAAC-----AGAGACGGACTGTGGC-CCTGTGCTTC----CACCGAGGGAGACACT----AGTTGACA
       .::..:     :::::: :.    : ::  :::: :  :    ::: :.:: . .::::    ::: :.::
       ATGTGCCACCCAGAGACTGGGGCTTGCGTCTGTCCCCCAGGACACAGTGGTGCGCACTGCAAAGTGGGCA
         2310      2320      2330      2340      2350      2360      2370

       1640      1650      1660          1670        1680          1690
inputs ---AAGTGTCTAAC-CCTCTTTTCCAACC-CAC---TGCTC---AAGTCCCTGTGGAC----ATAAGC--
          :.:.::: ..: ::. ..: :: ::: :.:   :: ::   :: :: ::: : .:    ::..::
       GCCAGGAGTCCTTCACCATAATGCCCACCTCTCCTGTGATCCATAACTCACTGGGTGCCGTGATTGGCAT
         2380      2390      2400      2410      2420      2430      2440
```

Figure 35D

```
          1700                1710      1720                1730      1740
inputs TGGTGGGCAGAA------TGTTGTTGTACAAGTG----TGATTTTAG---ATCGATTTTTTTTAAAGT-
       :: .: ::.:..      ::: :: :  :..::.   ::.::.:.:   : ::: . :    ..::.:
       TGCAGTGCTGGGGACCCTTGTGGTGGCCCTGGTAGCACTGTTTATTGGCTACCGACACTGGCAAAAGGGC
         2450      2460      2470      2480      2490      2500      2510

    1750      1760      1770        1780      1790      1800      1810
inputs ATGTGTTGGGTAC-CTTTTCTGTG--TGTATGCTCAGGCAGGCTGTGTGTGTCTCTAGTTGGCTTTAGAG
       :.: .. . :..: :::  :.:::  : . .::.:.::  :.::: .::  :: : : :. :  .:.  .
       AAGGAACATGAGCACTTGGCAGTGGCTTACAGCACTGGGCGACTGGATGGCTC-CGATTACGTCATGCCA
         2520      2530      2540      2550      2560      2570      2580

       1820            1830        1840         1850        1860      1870
inputs GGAGTC------AGGTATAGGTTCTGCCTT--CTGCACT---TTCCA-TCT-TATCT-AGTAGTCAGCTT
       :..:::      :: :: ::  .::.: .:  :..: ::   :.::: .:. :.::: :::. ::. ::.
       GATGTCTCTCCGAGCTACAGTCACTACTATTCCAACCCTAGCTACCACACACTGTCTCAGTGTTCTCCTA
         2590      2600      2610      2620      2630      2640      2650

            1880      1890           1900            1910          1920
inputs -CCAAGCTTAACTAGTTAGAGCTCCA--C---CAGCAG-----CAG-GCCCTAACTAC----CTGCCTGC
        :: . :.   : ... :... ::::  :    ::::.:     :::  :::...:..:    : :::..
       ACCCTCCACCCCCTAACAAGATTCCAGGCAGTCAGCTGTTTGTCAGCTCCCAGGCATCTGAGCGGCCAAA
         2660      2670      2680      2690      2700      2710      2720

       1930            1940       1950      1960       1970
inputs CCTTCACC-----C-AGTAATCCTC-CATGTCTTTGCTCAGA-GGATTGCTCC-CCGA----CTCT----
       :   . :::     : ::..:.::.: : . .::   : :.::  :...::.::  ::.    : :.
       CAGAAACCATGGGCGAGATAACCACGCCACACTGCCCGCTGACTGGAAGCACCGACGGGAGTCCCATGAC
         2730      2740      2750      2760      2770      2780      2790

    1980      1990          2000          2010            2020
inputs GGTGTTGTCCTCCTG----GTACGCCTTGAC----GGTCCTGCAGT--CT----CC-C-----TTTCCCG
       .:.: : :::::  :      .  :.::: :::     :::  .::.:: ::      :: :      :  ::::
       AGAGCTTTCCTCAGGCACCAGCCACCTGGACCGAAGGTATAGCTGTAGCTATGGCCACAGGAATGGCCCG
         2800      2810      2820      2830      2840      2850      2860

       2030      2040            2050      2060      2070        2080
inputs T---CTTGCT-TCATT------CTTTCCCAGAATGAAGGCTGTCTGCCACCCTACT-TCCCAGCCCAGGA
           :.: :: ::::.      : .:: :.::: .:.:.::.   :: : : :. : ::::.:  ::: .
       GGGCCATTCTGTCATAAAGGTCCCATCTCTGAAGAAGGACTAGGGGCAAGCGTTATGTCCCTGAGCAGTG
         2870      2880      2890      2900      2910      2920      2930

            2090        2100          2110      2120      2130      2140
inputs A-------TTGGCA--CATCTAAGTTCAGCC-----TTCCTAAGTTACCCGTTGAGTCCTGCTTGCCCTT
       :        :. ::.  :::: .::. :.:::      : :::..: .:::: .::.. :.:::.      .
       AGAACCCCTATGCGACCATCCGAGACCTGCCCGGCCTGCCTGGGGAACCCCGAGAAAGCAGCTATGTGGA
         2940      2950      2960      2970      2980      2990      3000

          2150         2160             2170      2180      2190      2200
inputs CACATAT-----TCCA-CAGAA-CACCCACC------CCACATCTGCTTCATAGCTACTCTCTTCTCCAC
       : ..:.      :::: :::.. :.::: ::       ::.:.::. :: :. ..:..     :. :. ::
       GATGAAAGGCCCTCCATCAGTGTCTCCCCCCAGGCAGCCTCTTCATCTCCGGGACAGGCAGCAGCAGCAA
         3010      3020      3030      3040      3050      3060      3070
```

Figure 35E

```
              2210      2220      2230      2240          2250      2260
inputs GTACCCACAGAAGGCAGAAGTGGTACCAGGCAAGAAGATGGGA---TTGTTGCATTTTGTTTTGTTTTTG
        :.:  .:. :..: .. :: :: :::..  .:: ::     . .   :::.  :.:..::.. .:: : ::
       CTGCAGTCTCAGAGAGACAGCGGCACCTAT-GAGCAGCCCACTCCCTTGAGCCGTAATGAAGAGTCTGTG
          3080      3090      3100       3110      3120      3130      3140

       2270      2280      2290      2300        2310      2320      2330
inputs AGACTCTGT-CTCACTATGTAGTCCTGGCTGGCCTG--GAACTCAAGAGCTCTGCCTGCCTCTGCCTCTT
       .: ::: .: : : :: : :  ::: :::  :::.   :. :.:.: ..:::  ::..  .:.:::.:.:
       GG-CTCCATGCCCCCTCT-TCCTCCGGGCCTGCCACCCGGCCACTATGACTCGCCCAAAAACAGCCACAT
            3150      3160      3170      3180      3190      3200      3210

              2340              2350        2360        2370      2380
inputs ----GAGTGCTGGGTTTA---------ACGGCT--CAGGGTCACATGCA---CAGCTCAAGCTGCACT--
           :.. .::. :. :.          :::::.  :.  .:::: : ::   :.:: : :: . :.::
       CCCTGGACACTATGACTTGCCTCCAGTACGGCATCCTCCATCACCTCCATCCCGGCGCCAGGACCGCTGA
              3220      3230      3240      3250      3260      3270      3280

                            2390            2400      2410      2420
inputs ----CCGA-------------TGTGCTT----TCCC---CTGTTGCTAGATTAGCGTCTGCCTCCC----
           ::..              .:::: :    .:::   :.: .::..:.  .: . :.::   ::
       GGAGCCAGCATGGTATGGGAGAGTGCCTGTGAACCCTGCCAGGAGCAGGGCCTGGACCAGCAGGCCATGA
              3290      3300      3310      3320      3330      3340      3350

                 2430              2440         2450        2460      2470
inputs ---------CCTAGTGGAG-------AGGCTGA---TCGC-CAGCT--CTCTGATGCAGGACTCTGGT--
                : :.:::.::       ::.:::    : :: :.:::  :.: :: :  ..:::.::.::
       ATAGACATACTTGGTGAAGTGAACGGAGACTGAGGATGGCTCTGCTTCCACCGAGG-GAGACACTAGTTG
              3360      3370      3380      3390      3400       3410

            2480         2490      2500                    2510
inputs GTTTAGGCTCA---CTCACTATTGGTTTCCTTGGCACAGG-----------GTAGTCA----CT------
       : ..:: ::.    ::: ::.:: .  ::.: ::.::.:             ::.:.::    ::
       GCAAAGTGTCTAACCTCCCTTTTCCAGCCCATTGCTCAAGTCCCCCAGGCTGTGGACATGAGCTGGTGGG
    3420      3430      3440      3450      3460      3470      3480

      2520        2530                                 2540      2550      2560
inputs CAA----TAAATGTTCC--TCT--------------------AAAAGCTGAAAAAAAAAAAAAAAAAAAGG
       ::.    :...::::   :::                    ::::. ..::::::::::::::::::::
       CAGAATGTTGTTGTTGAAGTCTGATTTTAGATTGATTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGG
    3490      3500      3510      3520      3530      3540      3550


inputs GCGGCCGC
       ::::::::
       GCGGCCGC
    3560
```

Figure 35F

```
               10        20        30        40        50        60        70
inputs MAPARAGFCPLLLLLLLGLWVAEIPVSAKPKGMTSSQWFKIQHMQPSPQACNSAMKNINKHTKRCKDLNT
       :.     .:  :::::: :: .  :. : :: .: ..:: :::.::::  :: :: .::. ...::. ::
       MV-----LCFPLLLLLLVLWGPVCPLHAWPKRLTKAHWFEIQHIQPSPLQCNRAMSGINNYAQHCKHQNT
                    10        20        30        40        50        60

               80        90       100       110       120       130       140
inputs FLHEPFSSVAATCQTPKIACKNGDKNCHQSHGPVSLTMCKLTSGKYPNCRYKEKRQNKSYVVACKPPQKK
       :::..: .:::.:.   :.:::   :::::  ::..: :.:::::::.::: .  : : ..::: ::::
       FLHDSFQNVAAVCDLLSIVCKNRRHNCHQSSKPVNMTDCRLTSGKYPQCRYSAAAQYKFFIVACDPPQKS
           70        80        90       100       110       120       130

              150
inputs DSQQFHLVPVHLDRVL
       :.. . ::::::: .:
       DPP-YKLVPVHLDSIL
          140       150
```

Figure 36

```
                 10        20        30        40        50        60        70
inputs GTCGACCCACGCGTCCGGCTCCCAGCCCACCCCCAAACAGACACAGCGTAGCCCGGGCCAGCTCTTAAGG
       .:                                                                  ::
       AT------------------------------------------------------------------GG

                 80        90       100       110       120       130       140
inputs AGTTCAGGAGTGAGAAGAGGCCCTCAGAGATCTGACAGCCTAGGAGTGCGTGGACACCACCTCAGCCCAC
       .:                                      :::    ::: :     .::.: ::.
       TG--------------------------------------CTA----TGCTT----TCCTCTTCT------
                                                       10          20

                150       160       170       180       190       200       210
inputs TGAGCAGGAGTCACAGCACGAAGACCAAGCGCAAAGCGACCCCTGCCCTCCATCCTGACTGCTCCTCCTA
       : ..:.:                                  ::::          ::. :       :::
       TTTACTG----------------------------------CTGC----------TGGTT-------CTA
           30                                                              40

                220       230       240       250       260       270       280
inputs AGAGAGATGGCACCGGCCAGAGCAGGATTCTGCCCCCTTCTGCTGCTTCTGCTGCTGGGGCTGTGGGTGG
       .:.:          :.:::::.:       :: :: ::::.  :::::            :::
       TGGG----------GACCAGTG--------TGTCCACTTCA--TGCTT------------GGC---------
                      50                  60          70

                290       300       310       320       330       340       350
inputs CAGAGATCCCAGTCAGTGCCAAGCCCAAGGGCATGACCTCATCACAGTGGTTTAAAATTCAGCACATGCA
       :..::    : :::.    ::   :::::: :          ::::  ::::::.::::::::::: ::.::
       CTAAG----C-GTCT----CA---CCAAGG-C----------TCAC--TGGTTTGAAATTCAGCATATACA
                 80                90                      100       110

                360       370       380       390       400       410       420
inputs GCCCAGCCCTCAAGCATGCAACTCAGCCATGAAAAACATTAACAAGCACACAAAACGGTGCAAAGACCTC
       ::: :: ::::.                :::              :.:::      ::.: ::::.::
       GCCAAGTCCTCT----------------CCA--------------ATGCA------ACAGGGCAATGA----
     120        130                                            140       150

                430       440       450       460       470       480       490
inputs AACACCTTCCTGCACGAGCCTTTCTCCAGTGTGGCCGCCACCTGCCAGACCCCCAAAATAGCCTGCAAGA
                                   ::::: : ::              ::.....:::
       --------------------------GTGGCATCAAC--------------AATTATGCC-------
                                         160                   170

                500       510       520       530       540       550       560
inputs ATGGCGATAAAAACTGCCACCAGAGCCACGGGCCCGTGTCCCTGACCATGTGTAAGCTCACCTCAGGGAA
                         :::   :::                         :::::::.    :::    :
       ------------------CAG---CAC-------------------------TGTAAGCA----TCA----A
                                                           180

                570       580       590       600       610       620       630
inputs GTATCCGAACTGCAGGTACAAAGAGAAGCGACAGAACAAGTCTTACGTAGTGGCCTGTAAGCCTCCCCAG
       ....:: ..::::: : ::                   :::::.:                      :::
       AATACCTTTCTGCATG-AC-------------------TCTTTC----------------------CAG
     190        200                          210

                640       650       660       670       680       690       700
inputs AAAAAGGACTCTCAGCAATTCCACCTGGTTCCTGTACACTTGGACAGAGTCCTTTAGGTTTCCAGACTGG
       ::...::      :.:::..:          ::::                       :.::: :.   :.:
```

Figure 37A

```
       AATGTGG-----CTGCTGT------------CTGT--------------------GATTTGCT--CAG-
        220          230                                          240

             710       720       730       740       750       760       770
inputs CTTGCTCTTTGGCTGACCTTCAATTCCCTCTCCAGGACTCCGCACCACTCCCCTACACCCAGAGCATTCT
             :.::: :::  :.. ::.::         :.:  :::: ::: ::      :::::::
       ------CATTGTCTG--CAA-AAATC---------GTCG--GCACAACTGCCA-----CCAGAGC-----
              250          260                 270            280

             780       790       800       810       820       830       840
inputs CTTCCCCTCATCTCTTGGGGCTGTTCCTGGTTCAGCCTCTGCTGGGAGGCTGAAGCTGACACTCTGGTGA
                   ::...:      ::::  :::.: :  .::    :.:::  :.:::.:.::
       ------------TCAAAG-------CCTG--TCAACAT-GACT----GACTG---CAGACTCACT-----
                     290                 300           310           320

             850       860       870       880       890       900       910
inputs GCTGAGCTCTAGAGGGATGCTTTTCATCTTTTTGTTGCTGTTTTCCCAGATGCTTATCCCCAAGAAACA
              ::..::.:                                       ::::::::.
       -------TCAGGAAAG---------------------------------------TATCCCCAG------
                  330

             920       930       940       950       960       970       980
inputs GCAAGCTCAGGTCTGTGGGTTCCCTGGTCTATGCCATTGCACATGTCTCCCCTGCCCCCTGGCATTAGGG
                                     ::::                          ::..::: :
       ------------------------------TGCC--------------------------GCTATAGTG
                                   340                                350

             990      1000      1010      1020      1030      1040      1050
inputs CAGCATGACAAGGAGAGGAAATAAATGGAAAGGGGGCATATGGGATTTGTGGACACAGCTGTTTCTGTTC
       :.::.                             ::                                  :
       CTGCT-----------------------------GC----------------------------------C
                                                                          360

            1060      1070      1080      1090      1100      1110      1120
inputs CTGAACTAGAAGTCTTCCCCAGCTCTGACGTGGCAGTGAGGTGACCTGAAGGAAAGAAAAATATAAATAA
       :.:.::.:..  :::::                                                  ::..
       CAGTACAAAT--TCTTC--------------------------------------------------ATTG
            370

            1130      1140      1150      1160      1170      1180      1190
inputs ATACCACTTCATATTTGTATAGAATCCTCTAATCCCTTGTGACATAGACTTGACAGGGATTGTATGCCTT
       .:.::.                :...::      :::            ::   :::
       TTGCCT----------------GTGACC------CCC------------CT---CAG--------------
     380.                      390

            1200      1210      1220      1230      1240      1250      1260
inputs CTTTATGGATGAGGAAATTAAGGTTTTAGAAAGCTTAATGAATTAAAGAGCTTGTCTAATTAGTTAGTAG
                                                 ::::::
       ------------------------------------------AAGAGC-------------------
                                               400

            1270      1280      1290      1300      1310      1320      1330
inputs CAGAACCTGGACTTGAACCTAGGTCTCCTTGCTCTAAATACAGTGTACCTTCTACTCTACCAGTTGCGCA
          .:::                ::                  ::       :::: :::::
       ---GACC----------------CC------------------CC-------CTACAAGTTG----
                            410                                 420

            1340      1350      1360      1370      1380      1390      1400
inputs AGAAAGAAGTCACTGTTACAGAGGCAAGCGGTGAACTAGGTAAGAGTTCACTCATGAAGAAACGAGTGCT
              . :: :::: :::               ::..:..::.::.::
       -------GTTC-CTGT-ACA---------------CTTAGATAGTATTCTCT------------------
                430                       440       450

            1410      1420      1430      1440      1450      1460      1470
```

Figure 37B

```
inputs CTGAAGAGCCAGTTACCCTGTGTTGGCTGCAATAAAGGTCATTACCTCTCTAGCCAAAAAAAAAAAAAAA
       ---------------------------------------------------------------------

                1480      1490
inputs AAAAAAAAAAAAAAAAAAAAAAAAAAA
                                ::
       -------------------------AA
```

Figure 37C

```
   240       250       260       270       280       290       300
AGGATTCTGCCCCCTTCTGCTGCTTCTGCTGCTGGGGCTGTGGGTGGCAGAGATCCCAGTCAGTGCCAAG
.: . : :::     : ::: ::  :.:::::::::::  ::.:::: . :::.:.  ::: :  .::: ..:
GGTGCTATGCTTTCCTCTTCTTTTACTGCTGCTGGTTCTATGGGGACCAGTGTGTCCACTTCATGCTTGG
   10        20        30        40        50        60        70

   310       320       330       340       350       360       370
CCCAAGGGCATGACCTCATCACAGTGGTTTAAAATTCAGCACATGCAGCCCAGCCCTCAAGCATGCAACT
:: ::: :   : :::. . :.:: ::::::.:::::::::: ::.::::: :: ::::.   :::::::.
CCTAAGCGTCTCACCAAGGCTCACTGGTTTGAAATTCAGCATATACAGCCAAGTCCTCTCCAATGCAACA
   80        90       100       110       120       130       140

   380       390       400       410       420       430       440
CAGCCATGAAAAACATTAACAAGCACACAAAACGGTGCAAAGACCTCAACACCTTCCTGCACGAGCCTTT
 .:: ::::...::: :::::  : .:  :.:. :: ::. : :. :: ::::: ::::: ::   ::::
GGGCAATGAGTGGCATCAACAATTATGCCCAGCACTGTAAGCATCAAAATACCTTTCTGCATGACTCTTT
   150       160       170       180       190       200       210

   450       460       470       480       490       500       510
CTCCAGTGTGGCCGCCACCTGCCAGACCCCCAAAATAGCCTGCAAGAAT-GGCGATAAAAACTGCCACCA
:   :.:::::: :: . :::  : .   : ::. ::.: ::::::.::: : ::. : :: :::::::::
CCAGAATGTGGCTGCTGTCTGTGATTTGCTCAGCATTGTCTGCAAAAATCGTCGGCACAA-CTGCCACCA
   220       230       240       250       260       270       280

   520       530       540       550       560       570       580
GAGCCACGGGCCCGTGTCCCTGACCATGTGTAAGCTCACCTCAGGGAAGTATCCGAACTGCAGGTACAAA
::::   ..::: :: . : :::: .. :: :..::::: ::::::.:::::::::  : ::: : :: :..
GAGCTCAAAGCCTGTCAACATGACTGACTGCAGACTCACTTCAGGAAAGTATCCCCAGTGCCGCTATAGT
   290       300       310       320       330       340       350

   590       600       610       620       630       640       650
G-AGAAGCGACAGAACAAGTCTTACGTAGTGGCCTGTAAGCCTCCCCAGAAAAAGGACTCTCAGCAATTC
: .: .::  :::.::::.:  :.:.:.:: :::::::.: :: :: :::::.:. ::: : :  :  :.:
GCTGCTGC-CCAGTACAAATTCTTCATTGTTGCCTGTGACCCCCCTCAGAAGAGCGACCCCCC-C--TAC
  360         370       380       390       400       410

   660       670       680
CACCTGGTTCCTGTACACTTGGACAGAGTCCTTTAG
 :  ::::::::::::::::.:: ::..: :: ::.
AAGTTGGTTCCTGTACACTTAGATAGTATTCTCTAA
420       430       440       450
```

43.4% identity in 477 aa overlap; score: 746

```
410          420          430         440          450        460
GGTGCAAAG---ACCTCAACACCTTC--CTGCACGAGCCTTTC--TCCAGTGTGGCCGCCACCTGCCAGA
::::::.:.:   .::::..:.  :.:  ::::. :. : .:      .::::::: ::.: .: ::: .:.
GGTGCTATGCTTTCCTCTTCTTTTACTGCTGCTGGTTCTATGGGGACCAGTGTGTCCACTTCATGCTTGG
    10        20        30        40        50        60        70
```

Figure 38A

```
470          480       490       500       510       520       530
 CC-------CCCAAAATAGCCTGCAAGAATGGCGATAAA-AACTGCCACCAGAGCCACGGGCCCGTGTCC
 ::        : ::  :..::  .:..:..::. ..: :. :. :.: .:::.. ::.:    :: .::
 CCTAAGCGTCTCACCAAGGCTCACTGGTTTGAAATTCAGCATATACAGCCAAGTCCTCT--CCAATGCAA
     80        90       100       110       120       130       140

     540       550       560       570       580       590       600
 CTGACCATGTGTAAGCTCACCTCAGGGAAGTATCCGAACTGCAGGTACAAAGAGAAGCGACAGAACAAGT
 :.:. ::.. . ..::. .: .::.  :.:   :: :.:.  . ...::. .:.:..: . . ..::.:.
 CAGGGCAATGAGTGGCA-TCAACAATTATG---CCCAGCACTGTAAGCATCAAAATACCTTTCTGCATGA
      150       160          170       180       190       200

     610       620       630       640        650       660
 CTTACGTAGTGGCCTGTAAGCCTCCCCAGAAAAAGGACT-CTCAGCAAT-TCCACCTGGTTCCTGTACAC
 ::  : :.      ::.:.:.:   :   :.:. ...:. : :::::::.: :: .: ....:: :   .:::
 CT--CTTT----CCAGAATGTGGCTGCTGTCTGTGATTTGCTCAGCATTGTCTGCAAAAATCGTCGGCAC
   210            220       230       240       250       260       270

670       680       690       700       710       720       730
 TTGGACAGAGTCCTTTAGGTTTCCAGACTGGCTTGCTCTTTGGCTGACCTTCAATTCCCTCTCCAGGA--
 .   ::.:   ::.   ::.   :: :..:    ::  :.   .::.:::: : ::.. : :.:: :::::
 A---ACTG---CCACCAGAGCTCAAAGC---CTGTCAACATGACTGAC-TGCAGA-CTCACTTCAGGAAA
               280       290          300       310       320

   740        750          760       770       780       790
 ----CTCC-GCACCACTCCC---CTACA-CCCAGAGCATTCTCTTCCCCTCATCTCTTGGGGCTGTTC-C
     : :: : .::.::         ::.:. ::::...::.. :::::    : .:   : :: :.:     : :
 GTATCCCCAGTGCCGCTATAGTGCTGCTGCCCAGTACAAATTCTTCA--TTGTTGCCTGTGACCCCCCTC
330       340       350       360       370        380       390

  800       810       820       830       840       850
 TG--GTTCAGCCTCTGCTGGGAGGCTGAAGCTGACACTCTGGTGAGCTGAGCTCTAG
 .:  :. :..:: :  ::. .::   :.. :::. ::.:: . ... :.. :::::.
 AGAAGAGCGACCCCCCCTACAAGTTGGTTCCTGT-ACACTTAGATAGTATTCTCTAA
 400       410       420       430        440       450
```

```
46.5% identity in 488 aa overlap; score:  709

      440       450       460         470        480          490
 TGCACGAGCCTTTCTCCAGTGTGGCCGCCACCTG--CCA-GACCCCCAAAATAGCC--TGCAAGAATGGC
 :::. ..:: :: ::: . : :  : :: .:  :   : : :.    :::...:. ::   : ::.: .::::
 TGCT-ATGCTTTCCTCTTCTTTTACTGCTGCTGGTTCTATGGGGACCAGTGTGTCCACTTCATGCTTGGC
     10        20        30        40        50        60        70

  500       510       520       530       540       550          560
 GATAAAAACTGCCACCAGAGC-CACGGGCCCGTGTCCCTGACCATGTGTAAGCTCA-CCTCAGGGAAGTA
 ..:. ..::  :::::..:: ::: ::     :...   :.:   :::.:. :. :. . ::::
 CTAAGCGTCT--CACCAAGGCTCACTGGTTTGAAATTCAG--CATATACAGCCAAGTCCTC---------
     80        90       100       110        120       130

     570       580       590       600       610       620       630
 TCCGAA-CTGCAGGTACAAAGAGAAGCGACAGAACAAGTCTTACGTAGTGGCCTGTAAGCCTCCCCAGAA
 :::.:. :..:::: .:::.:::.::..: ::::: : .:.:   :: .    :::::::: ::       :
 TCCAATGCAACAGG-GCAATGAGTGGCATC--AACAATT-ATGCCCAGCA--CTGTAAGCATC------A
       140       150        160        170        180

     640       650       660       670       680       690       700
 AAAGGACTCTCAGCAATTCCACCTGGTTCCTGTACACTTGGACAGAGTCCTTTAGGTTTC-CAGACTGGC
 ::: . :: ::.:::. .:     ::   ::::.:.:     :   :.:.:     :: :.. :: : ::: : :
 AAATACCTTTCTGCATGACT--CT--TTCCAGAA---TGTGGCTGCTGTCTGTGATTTGCTCAGCATTGT
190       200        210            220       230       240       250
```

Figure 38B

```
laminin_EGF: domain 1 of 4, from 3 to 37: score -1.2, E = 0.59
                   *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrCdr.CkpG
                         +  G        d+         ++GqC+ C+ + +G+rC  +C +G
      mT272     3      ---HASG-------DP-------VHGQCR-CQAGWMGTRCHLpCPEG 31

                   yyglpsgdpgqgC<-*
                   ++g       + +C
      mT272    32  FWG------A-NC     37

EGF: domain 1 of 4, from 37 to 67: score 19.2, E = 0.1
                   *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
                      C+ ++ C+ngGtCv+  g              C+C+pG      + G+ C
      mT272    37     CSNTCTCKNGGTCVSENG---------NCVCAPG-----FRGPSC    67

                   *

      mT272         -   -

DSL: domain 1 of 1, from 10 to 67: score -21.2, E = 8.1
                   *->WstdkhiggrtslGfnleyrirvtCdenYYGegCnkFCrPrdDafgH
                           +   + +       + r +  C e   G+ C++ C        +g+
      mT272    10     --HGQCRCQAG----WMGTRCHLPCPEGFWGANCSNTCTCK---NGG 47

                   ytCdenGnklCleGWkGeyC<-*
                     +enGn  C++G +G+ C
      mT272    48  TCVSENGNCVCAPGFRGPSC     67

laminin_EGF: domain 2 of 4, from 41 to 80: score -1.5, E = 0.63
                   *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrCdr.CkpG
                      C+C + G      tC s          e G C+ C p++ G+ C r+C pG
      mT272    41     CTCKNGG-----TCVS-------ENGNCV-CAPGFRGPSCQRpCPPG 74

                   yyglpsgdpgqgC<-*
                    y      + +  C
      mT272    75  RY-----GKR--C     80

EGF: domain 2 of 4, from 80 to 110: score 11.8, E = 1.9
                   *->CapnnpCsng.GtCvntpggssdnfggytCeCppGdyylsytGkrC<
                      C +    C+n++  C+++ g             tC C  G       +tG++C
      mT272    80     CVQC-KCNNNhSSCHPSDG---------TCSCLAG-----WTGPDC   110

                   -*

      mT272         -   -

laminin_EGF: domain 3 of 4, from 83 to 123: score 25.6, E = 0.0012
                   *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrC.drCkpG
                      C Cn++       ++C++          + G C+ C+ + tG++C++ C pG
      mT272    83     CKCNNNH----SSCHP-------SDGTCS-CLAGWTGPDCsEACPPG 117
```

Figure 34A

```
                yyglpsgdpgqgC<-*
                ++gl       C
mT272    118 HWGL-------KC     123

EGF: domain 3 of 4, from 123 to 153: score 27.3, E = 0.00036
                *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
                   C++++ C++gGtC++  g         +C+C+pG      +tG++C
mT272    123       CSQLCQCHHGGTCHPQDG---------SCICTPG-----WTGPNC    153

                *

mT272         -   -

laminin_EGF: domain 4 of 4, from 127 to 172: score -5.5, E = 1.4
                *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrC.drCkpG
                   C+C++ G     tC++          G C  C p+ tG++C + C p
mT272    127       CQCHHGG-----TCHP-------QDGSCI-CTPGWTGPNClEGCPPR 160

                yyglpsg.dpgqgC<-*
                 +g  ++++ + +C
mT272    161 MFG-VNCsQLC-QC     172

EGF: domain 4 of 4, from 166 to 196: score 6.5, E = 5.8
                *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
                   C++++ C+ g  C++  g         C+CppG      +G +C
mT272    166       CSQLCQCDLGEMCHPETG---------ACVCPPG-----HSGADC    196

                *

mT272         -   -

//
```

Figure 39B

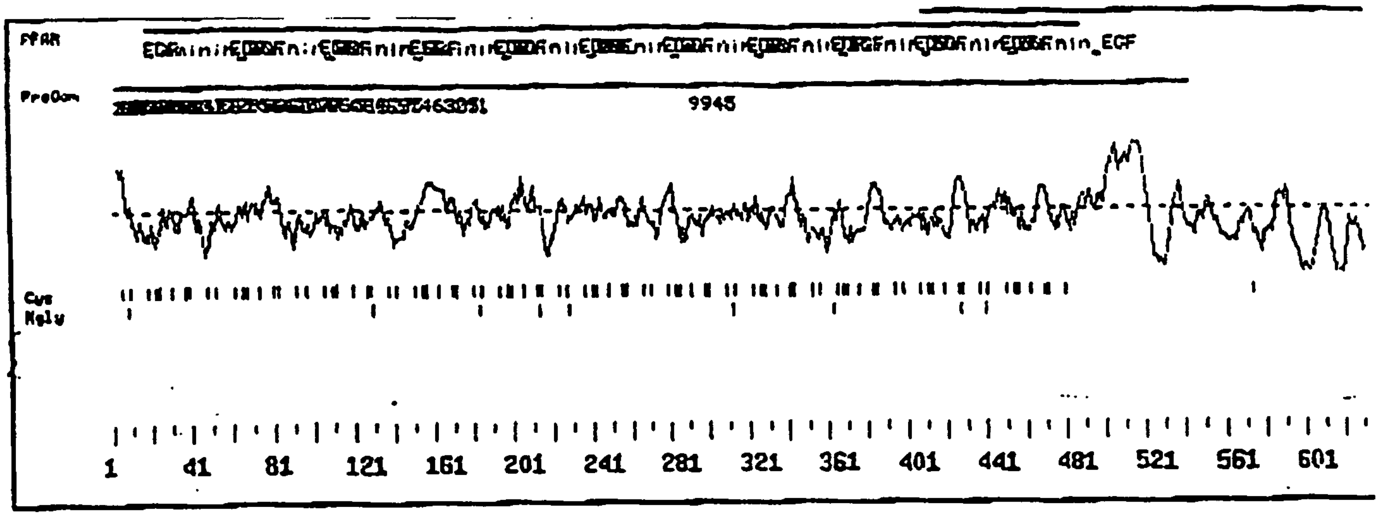
PFAM
PreDom
9945
Cys
Ngly
1
41
81
121
161
201
241
281
321
361
401
441
481
521
561
601

FIG. 40

```
                    *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
                       C++++ C+ngG C     g           +C+C+pG      y+G+rC
ratT272     18         ECRCHNGGLCDRFTG----------QCHCAPG ----YIGDRC    48

                    *

ratT272     -    -

laminin_EGF: domain 1 of 11, from 22 to 61: score 12.3, E = 0.038
                    *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrC.drCkpG
                       C C++ G       Cd+        +tGqC+ C p++ G+rC+++C  G
ratT272     22         CRCHNGG-----LCDR-------FTGQCH-CAPGYIGDRCrEECPVG 55

                    yyglpsgdpgqgC<-*
                     +g        q+C
ratT272     56      RFG-------QDC    61

EGF: domain 2 of 11, from 61 to 91: score 18.3, E = 0.18
                    *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
                       Ca+++ C  g++C +  g            C C +G      +tG+rC
ratT272     61         CAETCDCAPGARCFPANG---------ACLCEHG-----FTGDRC    91

                    *

ratT272     -    -

laminin_EGF: domain 2 of 11, from 65 to 105: score 4.0, E = 0.2
                    *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrCdr..Ckp
                       CdC p +      +C +         G+Cl C +++tG+rC ++ C +
ratT272     65         CDCAPGA-----RCFP-------ANGACL-CEHGFTGDRCTErlCPD 98

                    Gyyglpsgdpgqgc<-*
                    G ygl        +C
ratT272     99      GRYGL-------SC    105

EGF: domain 3 of 11, from 105 to 137: score 4.1, E = 9.6
                    *->CapnnpCsng..GtCvntpggssdnfggytCeCppGdyylsytGkrC
                       C++++ C+  ++  C++ +g           +C C+pG      ++G +C
ratT272     105        CQDPCTCDPEhsLSCHPMHG---------ECSCQPG-----WAGLHC 137

                    <-*

ratT272     -    -

laminin_EGF: domain 3 of 11, from 109 to 150: score 13.1, E = 0.032
                    *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrCdr.CkpG
                       C+C+p  sls    C++        ++G+C+ C+p+ +G +C+++C
ratT272     109        CTCDPEHSLS---CHP-------MHGECS-CQPGWAGLHCNEsCP-- 142

                    yyglpsgdpgqgC<-*
                      ++    + g gC
ratT272     143     --QD---THGAGC    150

EGF: domain 4 of 11, from 150 to 180: score 27.7, E = 0.00026
                    *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
                       C++++ C++gG+C+    g            C+C+pG      ytG++C
ratT272     150        CQEHCLCLHGGVCLADSG---------LCRCAPG-----YTGPHC    180
```

- FIGURE 41A

```
laminin_EGF: domain 4 of 11, from 154 to 193: score 8.4, E = 0.084
                        *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrC.drCkpG
                           C C +hG    + C           +G C+ C p++tG++C + C p+
   ratT272    154          CLC-LHG----GVCLA-------DSGLCR-CAPGYTGPHCaNLCPPN 187

                        yyglpsgdpgqgC<-*
                        +yg       +C
   ratT272    188 TYGI-------NC     193

EGF: domain 5 of 11, from 193 to 223: score 10.6, E = 2.5
                        *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
                           C++++ C n   C ++ g         tC+C++G      ++  +C
   ratT272    193          CSSHCSCENAIACSPVDG---------TCICKEG-----WQRGNC    223

                        *

   ratT272         -   -

laminin_EGF: domain 5 of 11, from 197 to 236: score 0.7, E = 0.4
                        *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrCdr.CkpG
                           C C ++        C +        + G C  Ck++ +  +C  +C pG
   ratT272    197          CSCENAI-----ACSP-------VDGTCI-CKEGWQRGNCSVpCPPG 230

                        yyglpsgdpgqgC<-*
                        ++g+      +C
   ratT272    231 TWGF-------SC     236

EGF: domain 6 of 11, from 236 to 266: score 11.8, E = 1.9
                        *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
                           C+   + C + G+C +  g           C+C+pG      + G +C
   ratT272    236          CNASCQCAHEGVCSPQTG---------ACTCTPG-----WRGVHC    266

                        *

   ratT272         -   -

laminin_EGF: domain 6 of 11, from 240 to 279: score -2.2, E = 0.73
                        *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrCdr.CkpG
                           C+C + G       C +        tG+C  C p+  G +C  +C  G
   ratT272    240          CQCAHEG-----VCSP-------QTGACT-CTPGWRGVHCQLpCPKG 273

                        yyglpsgdpgqgC<-*
                         +g       +gC
   ratT272    274 QFG-------EGC     279

DSL: domain 1 of 1, from 246 to 309: score -19.4, E = 5.2
                        *->WstdkhiggrtslGfnleyrirvtCdenYYGegCnkFCrPrdDafgH
                           +  ++++g+ t        +++     C +  +GegC+  C+       H
   ratT272    246          GVCSPQTGACTCTPGWRGVHCQLPCPKGQFGEGCASVCDCD-----H 287

                        yt.Cd.enGnklCleGWkGeyC<-*
                        + +Cd+ +G  +C +GW+G  C
   ratT272    288 SDgCDpVHGHCRCQAGWMGTRC     309

EGF: domain 7 of 11, from 279 to 309: score 7.0, E = 5.3
                        *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
                           Ca+ + C++   C +++g          +C+C+ G      + G rC
   ratT272    279          CASVCDCDHSDGCDPVHG---------HCRCQAG-----WMGTRC    309
```

FIG. 41B

```
laminin_EGF: domain 7 o. .1, from 283 to 322: score 12.7, . = 0.035
                     *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrCdr.CkpG
                        CdC+ h+    d Cd+        ++G+C+ C+ + +G+rC  +C +G
   ratT272   283        CDCD-HS----DGCDP-------VHGHCR-CQAGWMGTRCHLpCPEG 316

                     yyglpsgdpgqgC<-*
                     ++g       + +C
   ratT272   317 FWG------A-NC    322

EGF: domain 8 of 11, from 322 to 352: score 17.3, E = 0.38
                     *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
                        C+  + C+ngGtCv+  g            C+C+pG      + G+ C
   ratT272   322        CSNACTCKNGGTCVPENG---------NCVCAPG-----FRGPSC    352

                     *

   ratT272          -    -

laminin_EGF: domain 8 of 11, from 326 to 365: score -1.8, E = 0.67
                     *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrCdr.CkpG
                        C+C + G      tC +        e G C+ C p++ G+ C r+C pG
   ratT272   326        CTCKNGG-----TCVP-------ENGNCV-CAPGFRGPSCQRpCPPG 359

                     yyglpsgdpgqgC<-*
                      y      + +  C
   ratT272   360 RY-----GKR--C    365

EGF: domain 9 of 11, from 365 to 394: score 18.3, E = 0.18
                     *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
                        C p    C+n+  C+++ g            tC C  G      +tG++C
   ratT272   365        CVPC-KCNNHSSCHPSDG---------TCSCLAG-----WTGPDC    394

                     *

   ratT272          -    -

laminin_EGF: domain 9 of 11, from 368 to 407: score 24.0, E = 0.0034
                     *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrC.drCkpG
                        C Cn+h+      +C++        + G C+ C+ + tG++C++ C pG
   ratT272   368        CKCNNHS-----SCHP-------SDGTCS-CLAGWTGPDCsESCPPG 401

                     yyglpsgdpgqgC<-*
                     ++gl         C
   ratT272   402 HWGL-------KC    407

EGF: domain 10 of 11, from 407 to 437: score 24.0, E = 0.0035
                     *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
                        C++++ C++g+tC++  g            +C+C pG      +tG++C
   ratT272   407        CSQPCQCHHGATCHPQDG---------SCVCIPG-----WTGPNC    437

                     *

   ratT272          -    -

laminin_EGF: domain 10 of 11, from 411 to 450: score 6.5, E = 0.12
                     *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrCdrCkpGy
                        C+C++ +      tC++          G C+ C p+ tG++C +
   ratT272   411        CQCHHGA-----TCHP-------QDGSCV-CIPGWTGPNCSE----- 439

                     yglpsgdpgqgC<-*
                      g ps+++g++C
   ratT272   440 -GCPSRMFGVNC    450
```

FIG. 41C

```
EGF: domain 11 of 11, from 450 to 480: score 8.7, E = 3.7
                   *->CapnnpCsngGtCvntpggssdnfggytCeCppGdyylsytGkrC<-
                      C++++ C+ g  C++  g          C+CppG      +G +C
     ratT272   450    CSQLCQCDPGEMCHPETG---------ACVCPPG-----HSGAHC   480

                   *

     ratT272         -    -    .

laminin_EGF: domain 11 of 11, from 454 to 489: score -6.3, E = 1.7
                   *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrCdrCkpGy
                      C+C+p G    + C++       etG+C+ C p+ +G +C
     ratT272   454    CQCDP-G----EMCHP-------ETGACV-CPPGHSGAHC------K 481

                   yglpsgdpgqgC<-*
                    g  + ++
     ratT272   482 VGSQE-SFT---    489

//
```

FIG. 41D

SEQUENCE LISTING

<110> Millennium Pharmaceuticals, Inc.

<120> MEMBRANE-ASSOCIATED AND SECRETED PROTEINS AND USES THEREOF

<130> 7853-206-228

<150> 09/345,464
<151> 1999-06-30

<160> 148

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 3284
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1222)...(1944)

<400> 1

```
gtcgacccac gcgtccgtta tgtaactata cattttccca gaaattttag tatatgatat      60
gattttgttt tctttcatcc cttttcccaa gcagtttatt atgaaaattt tcaaacatac     120
agcaatgttg agaaaatttt acagtaaatg cctataccca ttacctaaat tttaccatta     180
acattttacc ctgctggcat tattgtgctt atccatctac gtatccctct ctcccttcat     240
tggtgtattt ctaagtaaat tgtaggcctc agtacacttc cttctgaatt cttcagcatg     300
cacaacagta ttatattcca tttttaaaag agcaattctt gatagattta tatagttttg     360
taaaatgttc atatagagct acaaatttta tctttttgtt tcttattgta tgtctagggt     420
cctgaagggg atgctggcat tgttgggata tcaggtccta aaggtcctat tggacacaga     480
ggaaacactg gtccccttgg cagagaaggt ataataggcc caacaggtag aactggaccc     540
agaggtgaaa agggctttag aggtgaaact ggtcctcaag gaccaagagg tcaaccaggg     600
cctccaggtc cacctggagc accaggccca agaaagcaaa tggatatcaa tgctgctatt     660
caagccttga ttgaatcaaa tactgcccta cagatggagg taacatatct ggttttattt     720
atattggcac tgtctctcaa tataccaatt aaacagagaa aatttttgga ggccaaaatg     780
tgacattatc tcaaagattg tatttaaaac agattgaaaa tgtgaaacca ttctcaagaa     840
caaagtaagt gattttggta taattaaaca gaaatatatg cgtaggatgt tttgtaagga     900
aaacatttaa atcaaaaatt tagtactgtt atttgtaagg aatttggtac tatccaagaa     960
agtagttaaa tgaggttagc catgtttctt aaaatgagat atatatatta tcactactca    1020
tttatttaaa ctctaatgat tcaatgtgta atttaaaaaa cataatacag tagacatagc    1080
aattcttatg ttagcttgaa aactaaactt gcaaatgtga atttaacctc tttaaaagat    1140
taaggttatt aaagcataca catatgccta tgcttaaata taaactgttc tttacattct    1200
actcacaact tactacacat a atg gaa aca cat tct tct cct gcc ttg gcc      1251
                        Met Glu Thr His Ser Ser Pro Ala Leu Ala
                         1               5                  10

cat gtt ggt cct cag gat ttt ttt gtt tat ata att ctt atg atg act      1299
His Val Gly Pro Gln Asp Phe Phe Val Tyr Ile Ile Leu Met Met Thr
                15                  20                  25

tgg cag agc tac cag aat act gaa gtg act tta att gac cac agt gaa      1347
Trp Gln Ser Tyr Gln Asn Thr Glu Val Thr Leu Ile Asp His Ser Glu
            30                  35                  40
```

```
gag ata ttc aaa acc ctg aac tac ctt agc aat tta ttg cac agc atc      1395
Glu Ile Phe Lys Thr Leu Asn Tyr Leu Ser Asn Leu Leu His Ser Ile
        45                  50                  55

aag aat cct ctt ggc aca cga gat aac cca gca cga atc tgc aaa gat      1443
Lys Asn Pro Leu Gly Thr Arg Asp Asn Pro Ala Arg Ile Cys Lys Asp
    60                  65                  70

tta ctt aac tgt gaa caa aaa gta tca gat gga aaa tac tgg att gac      1491
Leu Leu Asn Cys Glu Gln Lys Val Ser Asp Gly Lys Tyr Trp Ile Asp
 75                 80                  85                  90

cca aat ctt ggc tgt cct tca gat gcc att gag gtt ttc tgc aat ttc      1539
Pro Asn Leu Gly Cys Pro Ser Asp Ala Ile Glu Val Phe Cys Asn Phe
                 95                 100                 105

agt gct ggt ggc cag aca tgc tta cct cct gtt tct gta aca aag ttg      1587
Ser Ala Gly Gly Gln Thr Cys Leu Pro Pro Val Ser Val Thr Lys Leu
            110                 115                 120

gag ttt gga gtt ggg aaa gtc cag atg aac ttc ctt cat tta ctg agt      1635
Glu Phe Gly Val Gly Lys Val Gln Met Asn Phe Leu His Leu Leu Ser
        125                 130                 135

tcg gaa gcc acc cat atc atc acc att cac tgt cta aac acc cca agg      1683
Ser Glu Ala Thr His Ile Ile Thr Ile His Cys Leu Asn Thr Pro Arg
    140                 145                 150

tgg aca agc aca caa aca agt ggc cca gga ttg cct att ggt ttc aag      1731
Trp Thr Ser Thr Gln Thr Ser Gly Pro Gly Leu Pro Ile Gly Phe Lys
155                 160                 165                 170

gga tgg aat ggc cag att ttt aaa gta aac act cta ctt gaa cct aaa      1779
Gly Trp Asn Gly Gln Ile Phe Lys Val Asn Thr Leu Leu Glu Pro Lys
                175                 180                 185

gtg ctt tca gat gac tgc aag att caa gat ggc agc tgg cat aag gca      1827
Val Leu Ser Asp Asp Cys Lys Ile Gln Asp Gly Ser Trp His Lys Ala
            190                 195                 200

aca ttt ctt ttt cac acc cag gaa cct aat caa ctt cca gtg att gaa      1875
Thr Phe Leu Phe His Thr Gln Glu Pro Asn Gln Leu Pro Val Ile Glu
        205                 210                 215

gta caa aaa ctt cct cat ctc aaa act gaa cga aag tat tac att gac      1923
Val Gln Lys Leu Pro His Leu Lys Thr Glu Arg Lys Tyr Tyr Ile Asp
    220                 225                 230

agc agt tct gta tgc ttt ctg taaagtctct gaattagttc cgaattcagg         1974
Ser Ser Ser Val Cys Phe Leu
235                 240

ctgttggcca ggtaattgct gcagagggag aaataagaca gacagataca gtcattatga    2034
aatgcatgta ataaagcatt ggctaaatct taaagaatct caggaagaac agacttcctc    2094
ctaagaagga gaaaaggcat ttttaaagga ctatgattga taaagtattt aattctttta    2154
aaaattatat tcatctcagc tttcttagag aattccctag aactaaaaat ttataaatat    2214
ggaattcttc agggtatctt atatttttga ctgagtgcgt agtacccatt agacagctgg    2274
agatgcagag cactatggag caatactggc taatgcttcc agatgtgcac tgcttctgtc    2334
```

taaaaattac aagccacagt ctaatatgtc ttattttcca aaacactaag ctgtattcag 2394
gtccccgatg ggcatataca tcttagccgg tgatacacta cctcttacgt gttgcctctt 2454
tgtgttgctt ggtgctcttt cgaaaacaag gtgcttatgg ctttcataga ctatttcctt 2514
tttcatcttt gtcattcttt aaaagtgtat gtactggtta catcaagata tgttttggtt 2574
gttagtactt attttaattt gtttggtcac acacttaata acacatgaaa ctatttatgt 2634
gaagtccttg ttttatttta aaattctctt tgtgtatttg gaatcaaagc cagcacattg 2694
taacctgtgc ttgtacgcaa aagaattaga tttctttgtt tttgttttat tttttaaatt 2754
gttgtaaaaa ttattatagg ccagctacat ctagtagtag gtttggggta cagattgggg 2814
gttgtgccat actgttttta aagttcatga tcatctggaa tgatacttag tgtatatata 2874
ttttgtaaag ttttaattca gcaaattttt tgaaattgct gctgttttaa attataaaac 2934
ctttatattt ctgctttgta gaaattatat gttttgtagt attcattgat tttctttcac 2994
tgtacttaaa tttagtgtta gtactttaaa atttttaatt taccagtctt taaagcaaca 3054
tccagaaaaa aaaaagtctt ttcccattta aaataggctc agccagttca atgtcgcctt 3114
gttatcagag aaatattagt tcaatactga aagaaaaata ttatacctct tggtatctag 3174
aaaagcttgt tcatccatta taaatatatc tttagccaca gcaaaccaca cttaacctat 3234
ctataataaa aatgtgcttt aaataaaaaa aaaaaaaaaa agggcggccg 3284

<210> 2
<211> 241
<212> PRT
<213> Homo sapiens

<400> 2
Met Glu Thr His Ser Ser Pro Ala Leu Ala His Val Gly Pro Gln Asp
1 5 10 15
Phe Phe Val Tyr Ile Ile Leu Met Met Thr Trp Gln Ser Tyr Gln Asn
20 25 30
Thr Glu Val Thr Leu Ile Asp His Ser Glu Glu Ile Phe Lys Thr Leu
35 40 45
Asn Tyr Leu Ser Asn Leu Leu His Ser Ile Lys Asn Pro Leu Gly Thr
50 55 60
Arg Asp Asn Pro Ala Arg Ile Cys Lys Asp Leu Leu Asn Cys Glu Gln
65 70 75 80
Lys Val Ser Asp Gly Lys Tyr Trp Ile Asp Pro Asn Leu Gly Cys Pro
85 90 95
Ser Asp Ala Ile Glu Val Phe Cys Asn Phe Ser Ala Gly Gly Gln Thr
100 105 110
Cys Leu Pro Pro Val Ser Val Thr Lys Leu Glu Phe Gly Val Gly Lys
115 120 125
Val Gln Met Asn Phe Leu His Leu Leu Ser Ser Glu Ala Thr His Ile
130 135 140
Ile Thr Ile His Cys Leu Asn Thr Pro Arg Trp Thr Ser Thr Gln Thr
145 150 155 160
Ser Gly Pro Gly Leu Pro Ile Gly Phe Lys Gly Trp Asn Gly Gln Ile
165 170 175
Phe Lys Val Asn Thr Leu Leu Glu Pro Lys Val Leu Ser Asp Asp Cys
180 185 190
Lys Ile Gln Asp Gly Ser Trp His Lys Ala Thr Phe Leu Phe His Thr
195 200 205
Gln Glu Pro Asn Gln Leu Pro Val Ile Glu Val Gln Lys Leu Pro His
210 215 220
Leu Lys Thr Glu Arg Lys Tyr Tyr Ile Asp Ser Ser Ser Val Cys Phe
225 230 235 240
Leu

<210> 3
<211> 723

```
      &lt;212&gt; DNA
      &lt;213&gt; Homo sapiens

      &lt;400&gt; 3
atggaaacac attcttctcc tgccttggcc catgttggtc ctcaggattt ttttgtttat      60
ataattctta tgatgacttg gcagagctac cagaatactg aagtgacttt aattgaccac     120
agtgaagaga tattcaaaac cctgaactac cttagcaatt tattgcacag catcaagaat     180
cctcttggca cacgagataa cccagcacga atctgcaaag atttacttaa ctgtgaacaa     240
aaagtatcag atggaaaata ctggattgac ccaaatcttg gctgtccttc agatgccatt     300
gaggttttct gcaatttcag tgctggtggc cagacatgct tacctcctgt ttctgtaaca     360
aagttggagt ttggagttgg gaaagtccag atgaacttcc ttcatttact gagttcggaa     420
gccacccata tcatcaccat tcactgtcta aacaccccaa ggtggacaag cacacaaaca     480
agtggcccag gattgcctat tggtttcaag ggatggaatg gccagatttt taaagtaaac     540
actctacttg aacctaaagt gctttcagat gactgcaaga ttcaagatgg cagctggcat     600
aaggcaacat ttctttttca cacccaggaa cctaatcaac ttccagtgat tgaagtacaa     660
aaacttcctc atctcaaaac tgaacgaaag tattacattg acagcagttc tgtatgcttt     720
ctg                                                                   723

      &lt;210&gt; 4
      &lt;211&gt; 3169
      &lt;212&gt; DNA
      &lt;213&gt; Homo sapiens

      &lt;220&gt;
      &lt;221&gt; CDS
      &lt;222&gt; (57)...(1568)

      &lt;400&gt; 4
gtcgacccac gcgtccgcgc ccgctgagcc ccccgccgag gtccggacag gccgag atg      59
                                                              Met
                                                               1

acg ccg agc ccc ctg ttg ctg ctc ctg ctg ccg ccg ctg ctg ctg ggg       107
Thr Pro Ser Pro Leu Leu Leu Leu Leu Leu Pro Pro Leu Leu Leu Gly
            5                   10                  15

gcc ttc ccg ccg gcc gcc gcc gcc cga ggc ccc cca aag atg gcg gac       155
Ala Phe Pro Pro Ala Ala Ala Ala Arg Gly Pro Pro Lys Met Ala Asp
        20                  25                  30

aag gtg gtc cca cgg cag gtg gcc cgg ctg ggc cgc act gtg cgg ctg       203
Lys Val Val Pro Arg Gln Val Ala Arg Leu Gly Arg Thr Val Arg Leu
    35                  40                  45

cag tgc cca gtg gag ggg gac ccg ccg ccg ctg acc atg tgg acc aag       251
Gln Cys Pro Val Glu Gly Asp Pro Pro Pro Leu Thr Met Trp Thr Lys
 50                  55                  60                  65

gat ggc cgc acc atc cac agc ggc tgg agc cgc ttc cgc gtg ctg ccg       299
Asp Gly Arg Thr Ile His Ser Gly Trp Ser Arg Phe Arg Val Leu Pro
                70                  75                  80

cag ggg ctg aag gtg aag cag gtg gag cgg gag gat gcc ggc gtg tac       347
Gln Gly Leu Lys Val Lys Gln Val Glu Arg Glu Asp Ala Gly Val Tyr
            85                  90                  95

gtg tgc aag gcc acc aac ggc ttc ggc agc ctg agc gtc aac tac acc       395
Val Cys Lys Ala Thr Asn Gly Phe Gly Ser Leu Ser Val Asn Tyr Thr
```

```
        100                      105                      110

ctc gtc gtg ctg gat gac att agc cca ggg aag gag agc ctg ggg ccc     443
Leu Val Val Leu Asp Asp Ile Ser Pro Gly Lys Glu Ser Leu Gly Pro
    115                      120                      125

gac agc tcc tct ggg ggt caa gag gac ccc gcc agc cag cag tgg gca     491
Asp Ser Ser Ser Gly Gly Gln Glu Asp Pro Ala Ser Gln Gln Trp Ala
130                      135                      140                      145

cga ccg cgc ttc aca cag ccc tcc aag atg agg cgc cgg gtg atc gca     539
Arg Pro Arg Phe Thr Gln Pro Ser Lys Met Arg Arg Arg Val Ile Ala
                150                      155                      160

cgg ccc gtg ggt agc tcc gtg cgg ctc aag tgc gtg gcc agc ggg cac     587
Arg Pro Val Gly Ser Ser Val Arg Leu Lys Cys Val Ala Ser Gly His
            165                      170                      175

cct cgg ccc gac atc acg tgg atg aag gac gac cag gcc ttg acg cgc     635
Pro Arg Pro Asp Ile Thr Trp Met Lys Asp Asp Gln Ala Leu Thr Arg
        180                      185                      190

cca gag gcc gct gag ccc agg aag aag aag tgg aca ctg agc ctg aag     683
Pro Glu Ala Ala Glu Pro Arg Lys Lys Lys Trp Thr Leu Ser Leu Lys
    195                      200                      205

aac ctg cgg ccg gag gac agc ggc aaa tac acc tgc cgc gtg tcg aac     731
Asn Leu Arg Pro Glu Asp Ser Gly Lys Tyr Thr Cys Arg Val Ser Asn
210                      215                      220                      225

cgc gcg ggc gcc atc aac gcc acc tac aag gtg gat gtg atc cag cgg     779
Arg Ala Gly Ala Ile Asn Ala Thr Tyr Lys Val Asp Val Ile Gln Arg
                230                      235                      240

acc cgt tcc aag ccc gtg ctc aca ggc acg cac ccc gtg aac acg acg     827
Thr Arg Ser Lys Pro Val Leu Thr Gly Thr His Pro Val Asn Thr Thr
            245                      250                      255

gtg gac ttc ggg ggg acc acg tcc ttc cag tgc aag gtg cgc agc gac     875
Val Asp Phe Gly Gly Thr Thr Ser Phe Gln Cys Lys Val Arg Ser Asp
        260                      265                      270

gtg aag ccg gtg atc cag tgg ctg aag cgc gtg gag tac ggc gcc gag     923
Val Lys Pro Val Ile Gln Trp Leu Lys Arg Val Glu Tyr Gly Ala Glu
    275                      280                      285

ggc cgc cac aac tcc acc atc gat gtg ggc ggc cag aag ttt gtg gtg     971
Gly Arg His Asn Ser Thr Ile Asp Val Gly Gly Gln Lys Phe Val Val
290                      295                      300                      305

ctg ccc acg ggt gac gtg tgg tcg cgg ccc gac ggc tcc tac ctc aat     1019
Leu Pro Thr Gly Asp Val Trp Ser Arg Pro Asp Gly Ser Tyr Leu Asn
                310                      315                      320

aag ctg ctc atc acc cgt gcc cgc cag gac gat gcg ggc atg tac atc     1067
Lys Leu Leu Ile Thr Arg Ala Arg Gln Asp Asp Ala Gly Met Tyr Ile
            325                      330                      335
```

```
tgc ctt ggc gcc aac acc atg ggc tac agc ttc cgc agc gcc ttc ctc      1115
Cys Leu Gly Ala Asn Thr Met Gly Tyr Ser Phe Arg Ser Ala Phe Leu
        340                 345                 350

acc gtg ctg cca gac cca aaa ccg cca ggg cca cct gtg gcc tcc tcg      1163
Thr Val Leu Pro Asp Pro Lys Pro Pro Gly Pro Pro Val Ala Ser Ser
    355                 360                 365

tcc tcg gcc act agc ctg ccg tgg ccc gtg gtc atc ggc atc cca gcc      1211
Ser Ser Ala Thr Ser Leu Pro Trp Pro Val Val Ile Gly Ile Pro Ala
370                 375                 380                 385

ggc gct gtc ttc atc ctg ggc acc ctg ctc ctg tgg ctt tgc cag gcc      1259
Gly Ala Val Phe Ile Leu Gly Thr Leu Leu Leu Trp Leu Cys Gln Ala
                390                 395                 400

cag aag aag ccg tgc acc ccc gcg cct gcc cct ccc ctg cct ggg cac      1307
Gln Lys Lys Pro Cys Thr Pro Ala Pro Ala Pro Pro Leu Pro Gly His
            405                 410                 415

cgc ccg ccg ggg acg gcc cgc gac cgc agc gga gac aag gac ctt ccc      1355
Arg Pro Pro Gly Thr Ala Arg Asp Arg Ser Gly Asp Lys Asp Leu Pro
        420                 425                 430

tcg ttg gcc gcc ctc agc gct ggc cct ggt gtg ggg ctg tgt gag gag      1403
Ser Leu Ala Ala Leu Ser Ala Gly Pro Gly Val Gly Leu Cys Glu Glu
    435                 440                 445

cat ggg tct ccg gca gcc ccc cag cac tta ctg ggc cca ggc cca gtt      1451
His Gly Ser Pro Ala Ala Pro Gln His Leu Leu Gly Pro Gly Pro Val
450                 455                 460                 465

gct ggc cct aag ttg tac ccc aaa ctc tac aca gac atc cac aca cac      1499
Ala Gly Pro Lys Leu Tyr Pro Lys Leu Tyr Thr Asp Ile His Thr His
                470                 475                 480

aca cac aca cac tct cac aca cac tca cac gtg gag ggc aag gtc cac      1547
Thr His Thr His Ser His Thr His Ser His Val Glu Gly Lys Val His
            485                 490                 495

cag cac atc cac tat cag tgc tagacggcac cgtatctgca gtgggcacgg         1598
Gln His Ile His Tyr Gln Cys
        500

gggggccggc cagacaggca gactgggagg atggaggacg gagctgcaga cgaaggcagg   1658
ggacccatgg cgaggaggaa tggccagcac cccaggcagt ctgtgtgtga ggcatagccc   1718
ctggacacac acacacagac acacacactg cctggatgca tgtatgcaca cacatgcgcg   1778
cacacgtgct ccctgaaggc acacgtacgc acacacgcac atgcacagat atgccgcctg   1838
ggcacacaga taagctgccc aaatgcacgc acacgcacag agacatgcca gaacatacaa   1898
ggacatgctg cctgaacata cacacgcaca cccatgcgca gatgtgctgc ctggacacac   1958
acacacacac ggatatgctg tctggacgca cacacgtgca gatatggtat ccggacacac   2018
acgtgcacag atatgctgcc tggacacaca gataatgctg ccttgacaca cacatgcacg   2078
gatattgcct ggacacacac acacacacgt gtgcacagat atgctgtctg gacacgcaca   2138
cacatgcaga tatgctgcct ggacacacac ttccagacac acgtgcacag gcgcagatat   2198
gctgcctgga cacacgcaga tatgctgtct agtcacacac acacgcagac atgctgtccg   2258
gacacacaca cgcatgcaca gatatgctgt ccggacacac acacgcacgc agatatgctg   2318
cctggacaca cacacagata atgctgcctc aacactcaca cacgtgcaga tattgcctgg   2378
acacacacat gtgcacagat atgctgtctg gacatgcaca cacgtgcaga tatgctgtcc   2438
```

```
ggatacacac gcacgcacac atgcagatat gctgcctggg cacacacttc cggacacaca   2498
tgcacacaca ggtgcagata tgctgcctgg acacacgcag actgacgtgc ttttgggagg   2558
gtgtgccgtg aagcctgcag tacgtgtgcc gtgaggctca tagttgatga gggactttcc   2618
ctgctccacc gtcactcccc caactctgcc cgcctctgtc cccgcctcag tccccgcctc   2678
catccccgcc tctgtcccct ggccttggcg gctatttttg ccacctgcct tgggtgccca   2738
ggagtcccct actgctgtgg gctggggttg gggcacagc agccccaagc ctgagaggct   2798
ggagcccatg gctagtggct catccccact gcattctccc cctgacacag agaaggggcc   2858
ttggtattta tatttaagaa atgaagataa tattaataat gatggaagga agactgggtt   2918
gcagggactg tggtctctcc tggggcccgg gacccgcctg gtctttcagc catgctgatg   2978
accacacccc gtccaggcca gacaccaccc cccaccccac tgtcgtggtg gccccagatc   3038
tctgtaattt tatgtagagt ttgagctgaa gccccgtata tttaatttat tttgttaaac   3098
atgaaagtgc aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa   3158
agggcggccg c                                                        3169
```

<210> 5
<211> 504
<212> PRT
<213> Homo sapiens

<400> 5

```
Met Thr Pro Ser Pro Leu Leu Leu Leu Leu Leu Pro Pro Leu Leu Leu
 1               5                  10                  15
Gly Ala Phe Pro Pro Ala Ala Ala Ala Arg Gly Pro Pro Lys Met Ala
            20                  25                  30
Asp Lys Val Val Pro Arg Gln Val Ala Arg Leu Gly Arg Thr Val Arg
        35                  40                  45
Leu Gln Cys Pro Val Glu Gly Asp Pro Pro Pro Leu Thr Met Trp Thr
    50                  55                  60
Lys Asp Gly Arg Thr Ile His Ser Gly Trp Ser Arg Phe Arg Val Leu
65                  70                  75                  80
Pro Gln Gly Leu Lys Val Lys Gln Val Glu Arg Glu Asp Ala Gly Val
                85                  90                  95
Tyr Val Cys Lys Ala Thr Asn Gly Phe Gly Ser Leu Ser Val Asn Tyr
            100                 105                 110
Thr Leu Val Val Leu Asp Asp Ile Ser Pro Gly Lys Glu Ser Leu Gly
        115                 120                 125
Pro Asp Ser Ser Ser Gly Gly Gln Glu Asp Pro Ala Ser Gln Gln Trp
    130                 135                 140
Ala Arg Pro Arg Phe Thr Gln Pro Ser Lys Met Arg Arg Arg Val Ile
145                 150                 155                 160
Ala Arg Pro Val Gly Ser Ser Val Arg Leu Lys Cys Val Ala Ser Gly
                165                 170                 175
His Pro Arg Pro Asp Ile Thr Trp Met Lys Asp Asp Gln Ala Leu Thr
            180                 185                 190
Arg Pro Glu Ala Ala Glu Pro Arg Lys Lys Lys Trp Thr Leu Ser Leu
        195                 200                 205
Lys Asn Leu Arg Pro Glu Asp Ser Gly Lys Tyr Thr Cys Arg Val Ser
    210                 215                 220
Asn Arg Ala Gly Ala Ile Asn Ala Thr Tyr Lys Val Asp Val Ile Gln
225                 230                 235                 240
Arg Thr Arg Ser Lys Pro Val Leu Thr Gly Thr His Pro Val Asn Thr
                245                 250                 255
Thr Val Asp Phe Gly Gly Thr Thr Ser Phe Gln Cys Lys Val Arg Ser
            260                 265                 270
Asp Val Lys Pro Val Ile Gln Trp Leu Lys Arg Val Glu Tyr Gly Ala
        275                 280                 285
Glu Gly Arg His Asn Ser Thr Ile Asp Val Gly Gly Gln Lys Phe Val
    290                 295                 300
```

Val Leu Pro Thr Gly Asp Val Trp Ser Arg Pro Asp Gly Ser Tyr Leu
305 310 315 320
Asn Lys Leu Leu Ile Thr Arg Ala Arg Gln Asp Asp Ala Gly Met Tyr
325 330 335
Ile Cys Leu Gly Ala Asn Thr Met Gly Tyr Ser Phe Arg Ser Ala Phe
340 345 350
Leu Thr Val Leu Pro Asp Pro Lys Pro Pro Gly Pro Pro Val Ala Ser
355 360 365
Ser Ser Ser Ala Thr Ser Leu Pro Trp Pro Val Val Ile Gly Ile Pro
370 375 380
Ala Gly Ala Val Phe Ile Leu Gly Thr Leu Leu Leu Trp Leu Cys Gln
385 390 395 400
Ala Gln Lys Lys Pro Cys Thr Pro Ala Pro Ala Pro Pro Leu Pro Gly
405 410 415
His Arg Pro Pro Gly Thr Ala Arg Asp Arg Ser Gly Asp Lys Asp Leu
420 425 430
Pro Ser Leu Ala Ala Leu Ser Ala Gly Pro Gly Val Gly Leu Cys Glu
435 440 445
Glu His Gly Ser Pro Ala Ala Pro Gln His Leu Leu Gly Pro Gly Pro
450 455 460
Val Ala Gly Pro Lys Leu Tyr Pro Lys Leu Tyr Thr Asp Ile His Thr
465 470 475 480
His Thr His Thr His Ser His Thr His Ser His Val Glu Gly Lys Val
485 490 495
His Gln His Ile His Tyr Gln Cys
500

<210> 6
<211> 1512
<212> DNA
<213> Homo sapiens

<400> 6
atgacgccga gccccctgtt gctgctcctg ctgccgccgc tgctgctggg ggccttcccg 60
ccggccgccg ccgcccgagg ccccccaaag atggcggaca aggtggtccc acggcaggtg 120
gcccggctgg gccgcactgt gcggctgcag tgcccagtgg agggggaccc gccgccgctg 180
accatgtgga ccaaggatgg ccgcaccatc cacagcggct ggagccgctt ccgcgtgctg 240
ccgcaggggc tgaaggtgaa gcaggtggag cgggaggatg ccggcgtgta cgtgtgcaag 300
gccaccaacg gcttcggcag cctgagcgtc aactacaccc tcgtcgtgct ggatgacatt 360
agcccaggga aggagagcct ggggcccgac agctcctctg gggtcaaga ggaccccgcc 420
agccagcagt gggcacgacc gcgcttcaca cagccctcca agatgaggcg ccgggtgatc 480
gcacggcccg tgggtagctc cgtgcggctc aagtgcgtgg ccagcgggca ccctcggccc 540
gacatcacgt ggatgaagga cgaccaggcc ttgacgcgcc cagaggccgc tgagcccagg 600
aagaagaagt ggacactgag cctgaagaac ctgcggccgg aggacagcgg caaatacacc 660
tgccgcgtgt cgaaccgcgc gggcgccatc aacgccacct acaaggtgga tgtgatccag 720
cggacccgtt ccaagcccgt gctcacaggc acgcaccccg tgaacacgac ggtggacttc 780
ggggggacca cgtccttcca gtgcaaggtg cgcagcgacg tgaagccggt gatccagtgg 840
ctgaagcgcg tggagtacgg cgccgagggc cgccacaact ccaccatcga tgtgggcggc 900
cagaagtttg tggtgctgcc cacgggtgac gtgtggtcgc ggcccgacgg ctcctacctc 960
aataagctgc tcatcacccg tgcccgccag gacgatgcgg gcatgtacat ctgccttggc 1020
gccaacacca tgggctacag cttccgcagc gccttcctca ccgtgctgcc agacccaaaa 1080
ccgccagggc cacctgtggc ctcctcgtcc tcggccacta gcctgccgtg gcccgtggtc 1140
atcggcatcc cagccggcgc tgtcttcatc ctgggcaccc tgctcctgtg gctttgccag 1200
gcccagaaga agccgtgcac cccccgcgcct gcccctcccc tgcctgggca ccgcccgccg 1260
gggacggccc gcgaccgcag cggagacaag gaccttccct cgttggccgc cctcagcgct 1320
ggccctggtg tggggctgtg tgaggagcat gggtctccgg cagcccccca gcacttactg 1380
ggcccaggcc cagttgctgg ccctaagttg taccccaaac tctacacaga catccacaca 1440
cacacacaca cacactctca cacacactca cacgtggagg gcaaggtcca ccagcacatc 1500

```
cactatcagt gc                                                          1512

      <210> 7
      <211> 1074
      <212> DNA
      <213> Mus musculus

      <220>
      <221> CDS
      <222> (3)...(626)

      <221> modified_base
      <222> all "n" positions
      <223> n=a, c, g, or t

      <400> 7
ca cgc gtc cgg ccc acg ggt gat gtg tgg tca cgg cct gat ggc tcc          47
   Arg Val Arg Pro Thr Gly Asp Val Trp Ser Arg Pro Asp Gly Ser
    1               5                  10                  15

tac ctc aac aag ctg ctc atc tct cgg gcc cgc cag gat gat gct ggc         95
Tyr Leu Asn Lys Leu Leu Ile Ser Arg Ala Arg Gln Asp Asp Ala Gly
                20                  25                  30

atg tac atc tgc cta ggt gca aat acc atg ggc tac agt ttc cgt agc        143
Met Tyr Ile Cys Leu Gly Ala Asn Thr Met Gly Tyr Ser Phe Arg Ser
            35                  40                  45

gcc ttc ctc act gta tta cca gac ccc aaa cct cca ggg cct cct atg        191
Ala Phe Leu Thr Val Leu Pro Asp Pro Lys Pro Pro Gly Pro Pro Met
        50                  55                  60

gct tct tca tcg tca tcc aca agc ctg cca tgg cct gtg gtg atc ggc        239
Ala Ser Ser Ser Ser Ser Thr Ser Leu Pro Trp Pro Val Val Ile Gly
    65                  70                  75

atc cca gct ggt gct gtc ttc atc cta ggc act gtg ctg ctc tgg ctt        287
Ile Pro Ala Gly Ala Val Phe Ile Leu Gly Thr Val Leu Leu Trp Leu
80                  85                  90                  95

tgc cag acc aag aag aag cca tgt gcc cca gca tct aca ctt cct gtg        335
Cys Gln Thr Lys Lys Lys Pro Cys Ala Pro Ala Ser Thr Leu Pro Val
                100                 105                 110

cct ggg cat cgt ccc cca ggg aca tcc cga gaa cgc agt ggt gac aag        383
Pro Gly His Arg Pro Pro Gly Thr Ser Arg Glu Arg Ser Gly Asp Lys
            115                 120                 125

gac ctg ccc tca ttg gct gtg ggc ata tgt gag gag cat gga tcc gcc        431
Asp Leu Pro Ser Leu Ala Val Gly Ile Cys Glu Glu His Gly Ser Ala
        130                 135                 140

atg gcc ccc cag cac atc ctg gcc tct ggc tca act gct ggc ccc aag        479
Met Ala Pro Gln His Ile Leu Ala Ser Gly Ser Thr Ala Gly Pro Lys
    145                 150                 155

ctg tac ccc aag cta tac aca gat gtg cac aca cac aca cat aca cac        527
Leu Tyr Pro Lys Leu Tyr Thr Asp Val His Thr His Thr His Thr His
```

```
160                 165                 170                 175

acc tgc act cac acg ctc tca tgt tgg agg gca agg ttc atc aac acc       575
Thr Cys Thr His Thr Leu Ser Cys Trp Arg Ala Arg Phe Ile Asn Thr
                180                 185                 190

agc atg tcc act atc agt gct aaa tac agc gaa tct cca agc act gtg       623
Ser Met Ser Thr Ile Ser Ala Lys Tyr Ser Glu Ser Pro Ser Thr Val
            195                 200                 205

tcc tgaggtaggc atttgggggc caaggcaaca ggttgggaga attgagaaca            676
Ser


atggaggaag agtatcttag ggtgccttat ggtggacact cacaaacttg gccatataga    736
tgtatgtact accagatgaa cagccagcca gattcacaca cgcacatgtt taaacgtgta    796
aacgtgtgca caactgcaca cacaacctga gaaaccttca ggaggatttg tggtgtgact    856
ttgcagtgac atgtagcgat ggctagttga aggaatctcc ctcatgtctt agtggtcatg    916
gccacttccc cacccctgcc catctgtgtt cctgcctggc cttggtggtg cttccgtgtg    976
ccctgggttt tccaggaacc ctatcaacct gactggggtg agcagtgcag ccatgcntgg   1036
aggtttgagc caccctcccc ttgctagaga gaagggcn                           1074

      <210> 8
      <211> 208
      <212> PRT
      <213> Mus musculus

      <400> 8
Arg Val Arg Pro Thr Gly Asp Val Trp Ser Arg Pro Asp Gly Ser Tyr
 1               5                  10                  15
Leu Asn Lys Leu Leu Ile Ser Arg Ala Arg Gln Asp Asp Ala Gly Met
            20                  25                  30
Tyr Ile Cys Leu Gly Ala Asn Thr Met Gly Tyr Ser Phe Arg Ser Ala
        35                  40                  45
Phe Leu Thr Val Leu Pro Asp Pro Lys Pro Pro Gly Pro Pro Met Ala
    50                  55                  60
Ser Ser Ser Ser Ser Thr Ser Leu Pro Trp Pro Val Val Ile Gly Ile
65                  70                  75                  80
Pro Ala Gly Ala Val Phe Ile Leu Gly Thr Val Leu Leu Trp Leu Cys
                85                  90                  95
Gln Thr Lys Lys Lys Pro Cys Ala Pro Ala Ser Thr Leu Pro Val Pro
            100                 105                 110
Gly His Arg Pro Pro Gly Thr Ser Arg Glu Arg Ser Gly Asp Lys Asp
        115                 120                 125
Leu Pro Ser Leu Ala Val Gly Ile Cys Glu Glu His Gly Ser Ala Met
    130                 135                 140
Ala Pro Gln His Ile Leu Ala Ser Gly Ser Thr Ala Gly Pro Lys Leu
145                 150                 155                 160
Tyr Pro Lys Leu Tyr Thr Asp Val His Thr His Thr His Thr His Thr
                165                 170                 175
Cys Thr His Thr Leu Ser Cys Trp Arg Ala Arg Phe Ile Asn Thr Ser
            180                 185                 190
Met Ser Thr Ile Ser Ala Lys Tyr Ser Glu Ser Pro Ser Thr Val Ser
        195                 200                 205

      <210> 9
      <211> 624
      <212> DNA
```

```
<213> Mus musculus

<400> 9
cgcgtccggc ccacgggtga tgtgtggtca cggcctgatg gctcctacct caacaagctg    60
ctcatctctc gggcccgcca ggatgatgct ggcatgtaca tctgcctagg tgcaaatacc   120
atgggctaca gtttccgtag cgccttcctc actgtattac cagaccccaa acctccaggg   180
cctcctatgg cttcttcatc gtcatccaca agcctgccat ggcctgtggt gatcggcatc   240
ccagctggtg ctgtcttcat cctaggcact gtgctgctct ggctttgcca gaccaagaag   300
aagccatgtg ccccagcatc tacacttcct gtgcctgggc atcgtccccc agggacatcc   360
cgagaacgca gtggtgacaa ggacctgccc tcattggctg tgggcatatg tgaggagcat   420
ggatccgcca tggccccca gcacatcctg gcctctggct caactgctgg ccccaagctg   480
taccccaagc tatacacaga tgtgcacaca cacacacata cacacacctg cactcacacg   540
ctctcatgtt ggagggcaag gttcatcaac accagcatgt ccactatcag tgctaaatac   600
agcgaatctc caagcactgt gtcc                                           624

<210> 10
<211> 1423
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (31)...(444)

<400> 10
gtcgacccac gcgtccgccc acgcgtccgg atg cct gga ccc aga gtg tgg ggg      54
                                 Met Pro Gly Pro Arg Val Trp Gly
                                  1               5

aaa tat ctc tgg aga agc cct cac tcc aaa ggc tgt cca ggc gca atg     102
Lys Tyr Leu Trp Arg Ser Pro His Ser Lys Gly Cys Pro Gly Ala Met
    10                  15                  20

tgg tgg ctg ctt ctc tgg gga gtc ctc cag gct tgc cca acc cgg ggc     150
Trp Trp Leu Leu Leu Trp Gly Val Leu Gln Ala Cys Pro Thr Arg Gly
 25                  30                  35                  40

tcc gtc ctc ttg gcc caa gag cta ccc cag cag ctg aca tcc ccc ggg     198
Ser Val Leu Leu Ala Gln Glu Leu Pro Gln Gln Leu Thr Ser Pro Gly
                 45                  50                  55

tac cca gag ccg tat ggc aaa ggc caa gag agc agc acg gac atc aag     246
Tyr Pro Glu Pro Tyr Gly Lys Gly Gln Glu Ser Ser Thr Asp Ile Lys
             60                  65                  70

gct cca gag ggc ttt gct gtg agg ctc gtc ttc cag gac ttc gac ctg     294
Ala Pro Glu Gly Phe Ala Val Arg Leu Val Phe Gln Asp Phe Asp Leu
         75                  80                  85

gag ccg tcc cag gac tgt gca ggg gac tct gtc aca gtg agc tgg gga     342
Glu Pro Ser Gln Asp Cys Ala Gly Asp Ser Val Thr Val Ser Trp Gly
     90                  95                 100

tgg ggg ggg tcc cgc cag gac tgt ggc cag gga gat tcc cgg ggt tgt     390
Trp Gly Gly Ser Arg Gln Asp Cys Gly Gln Gly Asp Ser Arg Gly Cys
105                 110                 115                 120

ggg aag tgg cgg tgc cct gaa tcc ccc atc tgg agg agg gat gaa ttt     438
```

```
Gly Lys Trp Arg Cys Pro Glu Ser Pro Ile Trp Arg Arg Asp Glu Phe
                125                 130                 135

tcc atg taggggcagt cgggcttggc ttaccgggga gcagtggtgg accccaggac      494
Ser Met


acagcctccc accagcgcct ccggggctgc catctgggcc ccacagagca aagagggcag    554
caagcaggcc ctgcgtttgg aaggcttatg aatggacaca caaatcttgc aaatctatgg    614
agccaggggc agggacgcac atattggttg ttaaaaatat gtcatcatgt atttgttgag    674
tgcctgctct atcaggtgag gaagctggac acaaataata acaaaagatt aagtcaccgt    734
tcacacttac cttggaagag ctattacaaa acttctaacg ccaaagcctt attcagaata    794
aggacatttt aaaaacagta cttgatggag tgatgcaagc ttgcagtccc agcagtatag    854
tcaggagact gaggctggag gatcagaggg ctggagccca gggttcaagg ccagcctaag    914
caacatagca agaccccatc tcaaaaataa gtaaataata aataaaaata aaaagagcac    974
attatctttt gatttaaatt ttatttatat caaaatgaca taaatttttg aactttattt   1034
tttaatttta aaatttttaa ttattatgga tacataatag ttgtaagact ttttgttttt   1094
taattaaagt tttctaaggc tgggcgcagt agctcatgtc tgtagtccca gcactttggg   1154
aggctgaggc gaaagaagca cttgagccca ggaatttgag accagcctgg gcaacatagc   1214
aagaccccat ctctacaaaa aaatttaaaa attagccaag tgtggtggca cgcacctgtg   1274
gtcccagcta caagggacgc tgaagtgaga ggatcacttg agcctggaag gtagaggctg   1334
cagtgagctc tgatcatgac accgtactcc agcctgggtg acagagtgag accctgtctc   1394
caaaaaaaaa aaaaaaaaag ggcggccgc                                      1423

      <210> 11
      <211> 138
      <212> PRT
      <213> Homo sapiens

      <400> 11
Met Pro Gly Pro Arg Val Trp Gly Lys Tyr Leu Trp Arg Ser Pro His
 1               5                  10                  15
Ser Lys Gly Cys Pro Gly Ala Met Trp Trp Leu Leu Leu Trp Gly Val
            20                  25                  30
Leu Gln Ala Cys Pro Thr Arg Gly Ser Val Leu Leu Ala Gln Glu Leu
        35                  40                  45
Pro Gln Gln Leu Thr Ser Pro Gly Tyr Pro Glu Pro Tyr Gly Lys Gly
    50                  55                  60
Gln Glu Ser Ser Thr Asp Ile Lys Ala Pro Glu Gly Phe Ala Val Arg
65                  70                  75                  80
Leu Val Phe Gln Asp Phe Asp Leu Glu Pro Ser Gln Asp Cys Ala Gly
                85                  90                  95
Asp Ser Val Thr Val Ser Trp Gly Trp Gly Gly Ser Arg Gln Asp Cys
            100                 105                 110
Gly Gln Gly Asp Ser Arg Gly Cys Gly Lys Trp Arg Cys Pro Glu Ser
        115                 120                 125
Pro Ile Trp Arg Arg Asp Glu Phe Ser Met
    130                 135

      <210> 12
      <211> 414
      <212> DNA
      <213> Homo sapiens

      <400> 12
atgcctggac ccagagtgtg ggggaaatat ctctggagaa gccctcactc caaaggctgt     60
ccaggcgcaa tgtggtggct gcttctctgg ggagtcctcc aggcttgccc aacccggggc    120
tccgtcctct tggcccaaga gctaccccag cagctgacat cccccgggta cccagagccg    180
```

```
tatggcaaag gccaagagag cagcacggac atcaaggctc cagagggctt tgctgtgagg     240
ctcgtcttcc aggacttcga cctggagccg tcccaggact gtgcagggga ctctgtcaca     300
gtgagctggg gatgggggg gtcccgccag gactgtggcc agggagattc ccggggttgt      360
gggaagtggc ggtgccctga atccccatc tggaggaggg atgaattttc catg            414

      <210> 13
      <211> 5036
      <212> DNA
      <213> Homo sapiens

      <220>
      <221> CDS
      <222> (230)...(3379)

      <400> 13
gtcgacccac gcgtccgctc gaagcgggga ccctcgcccc gtcctcggct gtccagtcct     60
cctcctcgca gaccccggcg gttcctaccc caggccgcag gggagacggt gccccaaggc    120
aggcttcata tcctgaacgc tgggatcccc caggacattc cctggccccc aggccccagg    180
tcccaggccc cagggctgag ctgtgggcag gccccacctg gcctctgca atg tca ccg    238
                                                      Met Ser Pro
                                                       1

cct ctg tgt ccc ctc ctt ctc ctg gct gtg ggc ctg cgg ctg gct gga      286
Pro Leu Cys Pro Leu Leu Leu Leu Ala Val Gly Leu Arg Leu Ala Gly
     5                  10                  15

act ctc aac ccc agt gat ccc aat acc tgc agc ttc tgg gaa agc ttc      334
Thr Leu Asn Pro Ser Asp Pro Asn Thr Cys Ser Phe Trp Glu Ser Phe
 20                  25                  30                  35

act acc acc acc aag gag tcc cac tcc cgc ccc ttc agc ctg ctc ccc      382
Thr Thr Thr Thr Lys Glu Ser His Ser Arg Pro Phe Ser Leu Leu Pro
                 40                  45                  50

tca gag ccc tgc gag cgg ccc tgg gag ggc ccc cat act tgc ccc agc      430
Ser Glu Pro Cys Glu Arg Pro Trp Glu Gly Pro His Thr Cys Pro Ser
             55                  60                  65

cca caa act cag agg aaa ctc ctg gct tct agg gat tca ttc tgc atg      478
Pro Gln Thr Gln Arg Lys Leu Leu Ala Ser Arg Asp Ser Phe Cys Met
         70                  75                  80

gtc tgt gtc ggg gct gga gtg cag tgg cga gat cgt agt gca ctg caa      526
Val Cys Val Gly Ala Gly Val Gln Trp Arg Asp Arg Ser Ala Leu Gln
     85                  90                  95

cct caa aca ggg aat gcg ctt tct atg cgc cct cag ccc aga gtg ttg      574
Pro Gln Thr Gly Asn Ala Leu Ser Met Arg Pro Gln Pro Arg Val Leu
100                 105                 110                 115

agt ggt gcc cct tcc ctg gcc tcc cct ggc cac act gtg gtg gtg aag      622
Ser Gly Ala Pro Ser Leu Ala Ser Pro Gly His Thr Val Val Val Lys
                120                 125                 130

acg gac cac cgc cag cgc ctg cag tgc tgc cat ggc ttc tat gag agc      670
Thr Asp His Arg Gln Arg Leu Gln Cys Cys His Gly Phe Tyr Glu Ser
            135                 140                 145
```

```
agg ggg ttc tgt gtc ccg ctc tgt gcc cag gag tgt gtc cat ggc cgt      718
Arg Gly Phe Cys Val Pro Leu Cys Ala Gln Glu Cys Val His Gly Arg
        150                 155                 160

tgt gtg gca ccc aat cag tgc caa tgt gtg cca ggc tgg cgg ggc gac      766
Cys Val Ala Pro Asn Gln Cys Gln Cys Val Pro Gly Trp Arg Gly Asp
    165                 170                 175

gac tgt tcc agt gcc ccg aac tgc ctt cag ccc tgt acc cct ggc tac      814
Asp Cys Ser Ser Ala Pro Asn Cys Leu Gln Pro Cys Thr Pro Gly Tyr
180                 185                 190                 195

tat ggc cct gcc tgc cag ttc cgc tgc cag tgc cat ggg gca ccc tgc      862
Tyr Gly Pro Ala Cys Gln Phe Arg Cys Gln Cys His Gly Ala Pro Cys
                200                 205                 210

gat ccc cag act gga gcc tgc ttc tgc ccc gca gag aga act ggg ccc      910
Asp Pro Gln Thr Gly Ala Cys Phe Cys Pro Ala Glu Arg Thr Gly Pro
            215                 220                 225

agc tgt gac gtg tcc tgt tcc cag ggc act tct ggc ttc ttc tgc ccc      958
Ser Cys Asp Val Ser Cys Ser Gln Gly Thr Ser Gly Phe Phe Cys Pro
        230                 235                 240

agc acc cat cct tgc caa aat gga ggt gtc ttc caa acc cca cag ggc     1006
Ser Thr His Pro Cys Gln Asn Gly Gly Val Phe Gln Thr Pro Gln Gly
    245                 250                 255

tcc tgc agc tgc ccc cct ggc tgg atg ggc acc atc tgc tcc ctg ccc     1054
Ser Cys Ser Cys Pro Pro Gly Trp Met Gly Thr Ile Cys Ser Leu Pro
260                 265                 270                 275

tgc cca gag ggc ttt cac gga ccc aac tgc tcc cag gaa tgt cgc tgc     1102
Cys Pro Glu Gly Phe His Gly Pro Asn Cys Ser Gln Glu Cys Arg Cys
                280                 285                 290

cac aac ggc ggc ctc tgt gac cga ttc act ggg cag tgc cgc tgc gct     1150
His Asn Gly Gly Leu Cys Asp Arg Phe Thr Gly Gln Cys Arg Cys Ala
            295                 300                 305

ccg ggt tac act ggg gat cgg tgc cgg gag gag tgc ccg gtg ggc cgc     1198
Pro Gly Tyr Thr Gly Asp Arg Cys Arg Glu Glu Cys Pro Val Gly Arg
        310                 315                 320

ttt ggg cag gac tgt gct gag acg tgc gac tgc gcc ccg gac gcc cgt     1246
Phe Gly Gln Asp Cys Ala Glu Thr Cys Asp Cys Ala Pro Asp Ala Arg
    325                 330                 335

tgc ttc ccg gcc aac ggc gca tgt ctg tgc gaa cac ggc ttc act ggg     1294
Cys Phe Pro Ala Asn Gly Ala Cys Leu Cys Glu His Gly Phe Thr Gly
340                 345                 350                 355

gac cgc tgc acg gat cgc ctc tgc ccc gac ggc ttc tac ggt ctc agc     1342
Asp Arg Cys Thr Asp Arg Leu Cys Pro Asp Gly Phe Tyr Gly Leu Ser
                360                 365                 370

tgc cag gcc ccc tgc acc tgc gac cgg gag cac agc ctc agc tgc cac     1390
Cys Gln Ala Pro Cys Thr Cys Asp Arg Glu His Ser Leu Ser Cys His
```

```
            375                 380                 385

ccg atg aac ggg gag tgc tcc tgc ctg ccg ggc tgg gcg ggc ctc cac      1438
Pro Met Asn Gly Glu Cys Ser Cys Leu Pro Gly Trp Ala Gly Leu His
        390                 395                 400

tgc aac gag agc tgc ccg cag gac acg cat ggg cca ggg tgc cag gag      1486
Cys Asn Glu Ser Cys Pro Gln Asp Thr His Gly Pro Gly Cys Gln Glu
    405                 410                 415

cac tgt ctc tgc ctg cac ggt ggc gtc tgc cag gct acc agc ggc ctc      1534
His Cys Leu Cys Leu His Gly Gly Val Cys Gln Ala Thr Ser Gly Leu
420                 425                 430                 435

tgt cag tgc gcg ccg ggt tac acg ggc cct cac tgt gct agt ctt tgt      1582
Cys Gln Cys Ala Pro Gly Tyr Thr Gly Pro His Cys Ala Ser Leu Cys
                440                 445                 450

cct cct gac acc tac ggt gtc aac tgt tct gca cgc tgc tca tgt gaa      1630
Pro Pro Asp Thr Tyr Gly Val Asn Cys Ser Ala Arg Cys Ser Cys Glu
            455                 460                 465

aat gcc atc gcc tgc tca ccc atc gac ggc gag tgc gtc tgc aag gaa      1678
Asn Ala Ile Ala Cys Ser Pro Ile Asp Gly Glu Cys Val Cys Lys Glu
        470                 475                 480

ggt tgg cag cgt ggt aac tgc tct gtg ccc tgc cca ccc gga acc tgg      1726
Gly Trp Gln Arg Gly Asn Cys Ser Val Pro Cys Pro Pro Gly Thr Trp
    485                 490                 495

ggc ttc agt tgc aat gcc agc tgc cag tgt gcc cat gag gca gtc tgc      1774
Gly Phe Ser Cys Asn Ala Ser Cys Gln Cys Ala His Glu Ala Val Cys
500                 505                 510                 515

agc ccc caa act gga gcc tgt acc tgc acc cct ggg tgg cat ggg gcc      1822
Ser Pro Gln Thr Gly Ala Cys Thr Cys Thr Pro Gly Trp His Gly Ala
                520                 525                 530

cac tgc cag ctg ccc tgt ccg aag ggg cag ttt gga gaa ggt tgt gcc      1870
His Cys Gln Leu Pro Cys Pro Lys Gly Gln Phe Gly Glu Gly Cys Ala
            535                 540                 545

agt cgc tgt gac tgt gac cac tct gat ggc tgt gac cct gtt cat gga      1918
Ser Arg Cys Asp Cys Asp His Ser Asp Gly Cys Asp Pro Val His Gly
        550                 555                 560

cgc tgt cag tgc cag gct ggc tgg atg ggt gcc cgc tgc cac ctg tcc      1966
Arg Cys Gln Cys Gln Ala Gly Trp Met Gly Ala Arg Cys His Leu Ser
    565                 570                 575

tgc cct gag ggc tta tgg gga gtc aac tgt agc aac acc tgc acc tgc      2014
Cys Pro Glu Gly Leu Trp Gly Val Asn Cys Ser Asn Thr Cys Thr Cys
580                 585                 590                 595

aag aat ggg ggc acc tgt ctc cct gag aat ggc aac tgc gtg tgt gca      2062
Lys Asn Gly Gly Thr Cys Leu Pro Glu Asn Gly Asn Cys Val Cys Ala
                600                 605                 610
```

ccc gga ttc cgg ggc ccc tcc tgc cag aga tcc tgt cag cct ggc cgc 2110
Pro Gly Phe Arg Gly Pro Ser Cys Gln Arg Ser Cys Gln Pro Gly Arg
615 620 625

tat ggc aaa cgc tgt gtg ccc tgc aag tgc gct aac cac tcc ttc tgc 2158
Tyr Gly Lys Arg Cys Val Pro Cys Lys Cys Ala Asn His Ser Phe Cys
630 635 640

cac ccc tcg aac ggg acc tgc tac tgc ctg gct ggc tgg aca ggc ccc 2206
His Pro Ser Asn Gly Thr Cys Tyr Cys Leu Ala Gly Trp Thr Gly Pro
645 650 655

gac tgc tcc cag cca tgc cct cca gga cac tgg gga gaa aac tgt gcc 2254
Asp Cys Ser Gln Pro Cys Pro Pro Gly His Trp Gly Glu Asn Cys Ala
660 665 670 675

cag acc tgc caa tgt cac cat ggt ggg acc tgc cat ccc cag gat ggg 2302
Gln Thr Cys Gln Cys His His Gly Gly Thr Cys His Pro Gln Asp Gly
680 685 690

agc tgt atc tgc ccc cta ggc tgg act gga cac cac tgc tta gaa ggc 2350
Ser Cys Ile Cys Pro Leu Gly Trp Thr Gly His His Cys Leu Glu Gly
695 700 705

tgc cct ctg ggg aca ttt ggt gct aac tgc tcc cag cca tgc cag tgt 2398
Cys Pro Leu Gly Thr Phe Gly Ala Asn Cys Ser Gln Pro Cys Gln Cys
710 715 720

ggt cct gga gaa aag tgc cac cca gag act ggg gcc tgt gta tgt ccc 2446
Gly Pro Gly Glu Lys Cys His Pro Glu Thr Gly Ala Cys Val Cys Pro
725 730 735

cca ggg cac agt ggt gca cct tgc agg att gga atc cag gag ccc ttt 2494
Pro Gly His Ser Gly Ala Pro Cys Arg Ile Gly Ile Gln Glu Pro Phe
740 745 750 755

act gtg atg ccg acc act cca gta gcg tat aac tcg ctg ggt gca gtg 2542
Thr Val Met Pro Thr Thr Pro Val Ala Tyr Asn Ser Leu Gly Ala Val
760 765 770

att ggc att gca gtg ctg ggg tcc ctt gtg gta gcc ctg gtg gca ctg 2590
Ile Gly Ile Ala Val Leu Gly Ser Leu Val Val Ala Leu Val Ala Leu
775 780 785

ttc att ggc tat cgg cac tgg caa aaa ggc aag gag cac cac cac ctg 2638
Phe Ile Gly Tyr Arg His Trp Gln Lys Gly Lys Glu His His His Leu
790 795 800

gct gtg gct tac agc agc ggg cgc ctg gac ggc tcc gag tat gtc atg 2686
Ala Val Ala Tyr Ser Ser Gly Arg Leu Asp Gly Ser Glu Tyr Val Met
805 810 815

cca gat gtc cct ccg agc tac agt cac tac tac tcc aac ccc agc tac 2734
Pro Asp Val Pro Pro Ser Tyr Ser His Tyr Tyr Ser Asn Pro Ser Tyr
820 825 830 835

cac acc ctg tcg cag tgc tcc cca aac ccc cca ccc cct aac aag gtt 2782
His Thr Leu Ser Gln Cys Ser Pro Asn Pro Pro Pro Pro Asn Lys Val

```
                840                      845                      850

cca ggc ccg ctc ttt gcc agc ctg cag aac cct gag cgg cca ggt ggg      2830
Pro Gly Pro Leu Phe Ala Ser Leu Gln Asn Pro Glu Arg Pro Gly Gly
            855                     860                     865

gcc caa ggg cat gat aac cac acc acc ctg cct gct gac tgg aag cac      2878
Ala Gln Gly His Asp Asn His Thr Thr Leu Pro Ala Asp Trp Lys His
        870                     875                     880

cgc cgg gag ccc cct cca ggg cct ctg gac agg ggg agc agc cgc ctg      2926
Arg Arg Glu Pro Pro Pro Gly Pro Leu Asp Arg Gly Ser Ser Arg Leu
    885                     890                     895

gac cga agc tac agc tat agc tac agc aat ggc cca ggc cca ttc tac      2974
Asp Arg Ser Tyr Ser Tyr Ser Tyr Ser Asn Gly Pro Gly Pro Phe Tyr
900                     905                     910                     915

gat aaa ggg ctc atc tct gaa gag gag ctc ggg gcc agt gtg gct tcc      3022
Asp Lys Gly Leu Ile Ser Glu Glu Glu Leu Gly Ala Ser Val Ala Ser
                920                     925                     930

ctg agc agt gag aac cca tat gcc acc atc cgg gac ctg ccc agc ttg      3070
Leu Ser Ser Glu Asn Pro Tyr Ala Thr Ile Arg Asp Leu Pro Ser Leu
            935                     940                     945

cca ggg ggc ccc cgg gag agc agc tac atg gag atg aaa ggc cct ccc      3118
Pro Gly Gly Pro Arg Glu Ser Ser Tyr Met Glu Met Lys Gly Pro Pro
        950                     955                     960

tca gga tct gcc ccc agg cag cct cct cag ttt tgg gac agc cag agg      3166
Ser Gly Ser Ala Pro Arg Gln Pro Pro Gln Phe Trp Asp Ser Gln Arg
    965                     970                     975

cgg cgg caa ccc cag cca cag aga gac agt ggc acc tac gag cag ccc      3214
Arg Arg Gln Pro Gln Pro Gln Arg Asp Ser Gly Thr Tyr Glu Gln Pro
980                     985                     990                     995

agc ccc ctg atc cat gac cga gac tct gtg ggc tcc cag ccc cct ctg      3262
Ser Pro Leu Ile His Asp Arg Asp Ser Val Gly Ser Gln Pro Pro Leu
                1000                    1005                    1010

cct ccg ggc cta ccc ccc ggc cac tat gac tca ccc aag aac agc cac      3310
Pro Pro Gly Leu Pro Pro Gly His Tyr Asp Ser Pro Lys Asn Ser His
            1015                    1020                    1025

atc cct gga cat tat gac ttg cct cca gta cgg cat ccc cca tca cct      3358
Ile Pro Gly His Tyr Asp Leu Pro Pro Val Arg His Pro Pro Ser Pro
        1030                    1035                    1040

cca ctt cga cgc cag gac cgt tgaggagcca ggatggtatg gcagaggcca         3409
Pro Leu Arg Arg Gln Asp Arg
    1045                    1050

gcacacctgg ctgttgctgc tcaaggctgg ggacagagcc tagtgtaccc ctgccaggag   3469
cagggagtgg accggcaggc tgtgaacatg aacaacgctt aacagagcaa gtgatgggag   3529
ccttgttcct gggttctacc atgggagacg ctgatcagca ggatgcctgg ctccctttcc   3589
caacccactg ctcccaaggc ctccagggcc ctgtgtacat aaactggtgg gttggaagtt   3649
```

```
gctgggtaac tctgatttca gacatgcgtg tggggtacct tttctgtgca tgctcagcct   3709
gggctctgtg cgtgtgtgtg tttctgtgat tttagaaggg taccaggcag gttctgtcct   3769
agggcactta ccatttagta gggagatgga accaacccaa ttaactctag caatagcctc   3829
ctaactggcc tcctccattg attcagtgaa ccttccaatg catggctcat aatttcaaaa   3889
tacaggctgg ttagttactc cctacctgaa agccttcata ggtgcctctt tgctcttctg   3949
ccagtatcaa aacttttgaa ggccttaaag gccctgcttt gcctggccca tctgtctctc   4009
cagcctcacc ttgaactgtg ttcctgtcac tgcacgccag tcacaccggc ctctaggtcc   4069
tcctgtaggc cactcttctt tctggcacag ggacctgcac acctggagtg cccttcctcc   4129
cccactcgcc tgttcacccc tgcttttcct ttacacctcc tcctcaggga agtgcccacc   4189
ctccgtacat ctttcacagc cctgattgca gctgtgttca ctcaccaggt acctgcagaa   4249
ggcctacagg gtgccaggca cttctttaat gggttctttc tttatgtgat tatttgatta   4309
atctctgcct cccccactag actgtaagct ccctgaaggc aagaatcctg tgcttatgct   4369
caatattagc tctcccttgg cacagagtag gcactcaaca aatgctcccc aaaaggctga   4429
gtggctgact gaattaagta ccagtgacat gcagtaactg ctaagataga tgagccatct   4489
gtatgctctg acagttacag actgaataag ttggagactt ccctaaaggg tggcatttcc   4549
ccagggtaac aacgcagagc tcaggtgtgg gaaggtgcca ggggcagggg tgcagagggg   4609
ctgaggctga gggggtgca gaggctggag aaaggataac aggagagagt atacaggcat   4669
gccttgattt attgcacttc acaggtagca gaatttttaa agaaattgaa ggttttggga   4729
catatatgtg acagcaatag gttaagaaaa gcaaagcaga gaaattgaag atttgtgtca   4789
acactgcttt aagcaaatct gttggcacca tttttccaat agcatgtgcc cattttgggt   4849
ctctacattg cattttggta attgcttgca atatttcaag cattttcatt gttattatat   4909
gtgttatagt gatctgtgat cagtgatctt tgatatatta ttgtaattgt ttcggggcgc   4969
catgaaccgc acccatataa cacggtaaac ttaatcagca aaaaaaaaaa aaaaaaaagg   5029
gcggccg                                                            5036
```

```
      <210> 14
      <211> 1050
      <212> PRT
      <213> Homo sapiens

      <400> 14
Met Ser Pro Pro Leu Cys Pro Leu Leu Leu Leu Ala Val Gly Leu Arg
 1               5                  10                  15
Leu Ala Gly Thr Leu Asn Pro Ser Asp Pro Asn Thr Cys Ser Phe Trp
            20                  25                  30
Glu Ser Phe Thr Thr Thr Thr Lys Glu Ser His Ser Arg Pro Phe Ser
        35                  40                  45
Leu Leu Pro Ser Glu Pro Cys Glu Arg Pro Trp Glu Gly Pro His Thr
    50                  55                  60
Cys Pro Ser Pro Gln Thr Gln Arg Lys Leu Leu Ala Ser Arg Asp Ser
65                  70                  75                  80
Phe Cys Met Val Cys Val Gly Ala Gly Val Gln Trp Arg Asp Arg Ser
                85                  90                  95
Ala Leu Gln Pro Gln Thr Gly Asn Ala Leu Ser Met Arg Pro Gln Pro
            100                 105                 110
Arg Val Leu Ser Gly Ala Pro Ser Leu Ala Ser Pro Gly His Thr Val
        115                 120                 125
Val Val Lys Thr Asp His Arg Gln Arg Leu Gln Cys Cys His Gly Phe
    130                 135                 140
Tyr Glu Ser Arg Gly Phe Cys Val Pro Leu Cys Ala Gln Glu Cys Val
145                 150                 155                 160
His Gly Arg Cys Val Ala Pro Asn Gln Cys Gln Cys Val Pro Gly Trp
                165                 170                 175
Arg Gly Asp Asp Cys Ser Ser Ala Pro Asn Cys Leu Gln Pro Cys Thr
            180                 185                 190
Pro Gly Tyr Tyr Gly Pro Ala Cys Gln Phe Arg Cys Gln Cys His Gly
        195                 200                 205
Ala Pro Cys Asp Pro Gln Thr Gly Ala Cys Phe Cys Pro Ala Glu Arg
```

```
    210                 215                 220
Thr Gly Pro Ser Cys Asp Val Ser Cys Ser Gln Gly Thr Ser Gly Phe
225                 230                 235                 240
Phe Cys Pro Ser Thr His Pro Cys Gln Asn Gly Gly Val Phe Gln Thr
                245                 250                 255
Pro Gln Gly Ser Cys Ser Cys Pro Pro Gly Trp Met Gly Thr Ile Cys
            260                 265                 270
Ser Leu Pro Cys Pro Glu Gly Phe His Gly Pro Asn Cys Ser Gln Glu
        275                 280                 285
Cys Arg Cys His Asn Gly Gly Leu Cys Asp Arg Phe Thr Gly Gln Cys
    290                 295                 300
Arg Cys Ala Pro Gly Tyr Thr Gly Asp Arg Cys Arg Glu Glu Cys Pro
305                 310                 315                 320
Val Gly Arg Phe Gly Gln Asp Cys Ala Glu Thr Cys Asp Cys Ala Pro
                325                 330                 335
Asp Ala Arg Cys Phe Pro Ala Asn Gly Ala Cys Leu Cys Glu His Gly
            340                 345                 350
Phe Thr Gly Asp Arg Cys Thr Asp Arg Leu Cys Pro Asp Gly Phe Tyr
        355                 360                 365
Gly Leu Ser Cys Gln Ala Pro Cys Thr Cys Asp Arg Glu His Ser Leu
    370                 375                 380
Ser Cys His Pro Met Asn Gly Glu Cys Ser Cys Leu Pro Gly Trp Ala
385                 390                 395                 400
Gly Leu His Cys Asn Glu Ser Cys Pro Gln Asp Thr His Gly Pro Gly
                405                 410                 415
Cys Gln Glu His Cys Leu Cys Leu His Gly Gly Val Cys Gln Ala Thr
            420                 425                 430
Ser Gly Leu Cys Gln Cys Ala Pro Gly Tyr Thr Gly Pro His Cys Ala
        435                 440                 445
Ser Leu Cys Pro Pro Asp Thr Tyr Gly Val Asn Cys Ser Ala Arg Cys
    450                 455                 460
Ser Cys Glu Asn Ala Ile Ala Cys Ser Pro Ile Asp Gly Glu Cys Val
465                 470                 475                 480
Cys Lys Glu Gly Trp Gln Arg Gly Asn Cys Ser Val Pro Cys Pro Pro
                485                 490                 495
Gly Thr Trp Gly Phe Ser Cys Asn Ala Ser Cys Gln Cys Ala His Glu
            500                 505                 510
Ala Val Cys Ser Pro Gln Thr Gly Ala Cys Thr Cys Thr Pro Gly Trp
        515                 520                 525
His Gly Ala His Cys Gln Leu Pro Cys Pro Lys Gly Gln Phe Gly Glu
    530                 535                 540
Gly Cys Ala Ser Arg Cys Asp Cys Asp His Ser Asp Gly Cys Asp Pro
545                 550                 555                 560
Val His Gly Arg Cys Gln Cys Gln Ala Gly Trp Met Gly Ala Arg Cys
                565                 570                 575
His Leu Ser Cys Pro Glu Gly Leu Trp Gly Val Asn Cys Ser Asn Thr
            580                 585                 590
Cys Thr Cys Lys Asn Gly Gly Thr Cys Leu Pro Glu Asn Gly Asn Cys
        595                 600                 605
Val Cys Ala Pro Gly Phe Arg Gly Pro Ser Cys Gln Arg Ser Cys Gln
    610                 615                 620
Pro Gly Arg Tyr Gly Lys Arg Cys Val Pro Cys Lys Cys Ala Asn His
625                 630                 635                 640
Ser Phe Cys His Pro Ser Asn Gly Thr Cys Tyr Cys Leu Ala Gly Trp
                645                 650                 655
Thr Gly Pro Asp Cys Ser Gln Pro Cys Pro Pro Gly His Trp Gly Glu
            660                 665                 670
Asn Cys Ala Gln Thr Cys Gln Cys His His Gly Gly Thr Cys His Pro
```

```
        675                 680                 685
Gln Asp Gly Ser Cys Ile Cys Pro Leu Gly Trp Thr Gly His His Cys
    690                 695                 700
Leu Glu Gly Cys Pro Leu Gly Thr Phe Gly Ala Asn Cys Ser Gln Pro
705                 710                 715                 720
Cys Gln Cys Gly Pro Gly Glu Lys Cys His Pro Glu Thr Gly Ala Cys
                725                 730                 735
Val Cys Pro Pro Gly His Ser Gly Ala Pro Cys Arg Ile Gly Ile Gln
            740                 745                 750
Glu Pro Phe Thr Val Met Pro Thr Thr Pro Val Ala Tyr Asn Ser Leu
        755                 760                 765
Gly Ala Val Ile Gly Ile Ala Val Leu Gly Ser Leu Val Val Ala Leu
    770                 775                 780
Val Ala Leu Phe Ile Gly Tyr Arg His Trp Gln Lys Gly Lys Glu His
785                 790                 795                 800
His His Leu Ala Val Ala Tyr Ser Ser Gly Arg Leu Asp Gly Ser Glu
                805                 810                 815
Tyr Val Met Pro Asp Val Pro Pro Ser Tyr Ser His Tyr Tyr Ser Asn
            820                 825                 830
Pro Ser Tyr His Thr Leu Ser Gln Cys Ser Pro Asn Pro Pro Pro Pro
        835                 840                 845
Asn Lys Val Pro Gly Pro Leu Phe Ala Ser Leu Gln Asn Pro Glu Arg
    850                 855                 860
Pro Gly Gly Ala Gln Gly His Asp Asn His Thr Thr Leu Pro Ala Asp
865                 870                 875                 880
Trp Lys His Arg Arg Glu Pro Pro Pro Gly Pro Leu Asp Arg Gly Ser
                885                 890                 895
Ser Arg Leu Asp Arg Ser Tyr Ser Tyr Ser Tyr Ser Asn Gly Pro Gly
            900                 905                 910
Pro Phe Tyr Asp Lys Gly Leu Ile Ser Glu Glu Glu Leu Gly Ala Ser
        915                 920                 925
Val Ala Ser Leu Ser Ser Glu Asn Pro Tyr Ala Thr Ile Arg Asp Leu
    930                 935                 940
Pro Ser Leu Pro Gly Gly Pro Arg Glu Ser Ser Tyr Met Glu Met Lys
945                 950                 955                 960
Gly Pro Pro Ser Gly Ser Ala Pro Arg Gln Pro Pro Gln Phe Trp Asp
                965                 970                 975
Ser Gln Arg Arg Arg Gln Pro Gln Pro Gln Arg Asp Ser Gly Thr Tyr
            980                 985                 990
Glu Gln Pro Ser Pro Leu Ile His Asp Arg Asp Ser Val Gly Ser Gln
        995                 1000                1005
Pro Pro Leu Pro Pro Gly Leu Pro Pro Gly His Tyr Asp Ser Pro Lys
    1010                1015                1020
Asn Ser His Ile Pro Gly His Tyr Asp Leu Pro Pro Val Arg His Pro
1025                1030                1035                1040
Pro Ser Pro Pro Leu Arg Arg Gln Asp Arg
                1045                1050
```

```
<210> 15
<211> 3150
<212> DNA
<213> Homo sapiens

<400> 15
atgtcaccgc ctctgtgtcc cctccttctc ctggctgtgg gcctgcggct ggctggaact     60
ctcaacccca gtgatcccaa tacctgcagc ttctgggaaa gcttcactac caccaccaag    120
gagtcccact cccgcccctt cagcctgctc ccctcagagc cctgcgagcg gccctgggag    180
ggcccccata cttgccccag cccacaaact cagaggaaac tcctggcttc tagggattca    240
```

```
ttctgcatgg tctgtgtcgg ggctggagtg cagtggcgag atcgtagtgc actgcaacct    300
caaacaggga atgcgctttc tatgcgccct cagcccagag tgttgagtgg tgccccttcc    360
ctggcctccc ctggccacac tgtggtggtg aagacggacc accgccagcg cctgcagtgc    420
tgccatggct tctatgagag caggggggttc tgtgtcccgc tctgtgccca ggagtgtgtc    480
catggccgtt gtgtggcacc caatcagtgc caatgtgtgc caggctggcg gggcgacgac    540
tgttccagtg ccccgaactg ccttcagccc tgtacccctg gctactatgg ccctgcctgc    600
cagttccgct gccagtgcca tggggcaccc tgcgatcccc agactggagc ctgcttctgc    660
cccgcagaga gaactgggcc cagctgtgac gtgtcctgtt cccagggcac ttctggcttc    720
ttctgcccca gcacccatcc ttgccaaaat ggaggtgtct tccaaacccc acagggctcc    780
tgcagctgcc cccctggctg gatgggcacc atctgctccc tgccctgccc agagggcttt    840
cacggaccca actgctccca ggaatgtcgc tgccacaacg gcggcctctg tgaccgattc    900
actgggcagt gccgctgcgc tccgggttac actggggatc ggtgccggga ggagtgcccg    960
gtgggccgct ttgggcagga ctgtgctgag acgtgcgact gcgccccgga cgcccgttgc   1020
ttcccggcca acggcgcatg tctgtgcgaa cacggcttca ctggggaccg ctgcacggat   1080
cgcctctgcc ccgacggctt ctacggtctc agctgccagg cccccctgcac ctgcgaccgg   1140
gagcacagcc tcagctgcca cccgatgaac ggggagtgct cctgcctgcc gggctgggcg   1200
ggcctccact gcaacgagag ctgcccgcag gacacgcatg ggccagggtg ccaggagcac   1260
tgtctctgcc tgcacggtgg cgtctgccag gctaccagcg gcctctgtca gtgcgcgccg   1320
ggttacacgg gccctcactg tgctagtctt tgtcctcctg acacctacgg tgtcaactgt   1380
tctgcacgct gctcatgtga aaatgccatc gcctgctcac ccatcgacgg cgagtgcgtc   1440
tgcaaggaag gttggcagcg tggtaactgc tctgtgccct gcccacccgg aacctggggc   1500
ttcagttgca atgccagctg ccagtgtgcc catgaggcag tctgcagccc ccaaactgga   1560
gcctgtacct gcacccctgg gtggcatggg gcccactgcc agctgccctg tccgaagggg   1620
cagtttggag aaggttgtgc cagtcgctgt gactgtgacc actctgatgg ctgtgaccct   1680
gttcatggac gctgtcagtg ccaggctggc tggatgggtg cccgctgcca cctgtcctgc   1740
cctgagggct tatggggagt caactgtagc aacacctgca cctgcaagaa tgggggcacc   1800
tgtctccctg agaatggcaa ctgcgtgtgt gcacccggat tccggggccc ctcctgccag   1860
agatcctgtc agcctggccg ctatggcaaa cgctgtgtgc cctgcaagtg cgctaaccac   1920
tccttctgcc acccctcgaa cgggacctgc tactgcctgg ctggctggac aggccccgac   1980
tgctcccagc catgccctcc aggacactgg ggagaaaact gtgcccagac ctgccaatgt   2040
caccatggtg ggacctgcca tccccaggat gggagctgta tctgccccct aggctggact   2100
ggacaccact gcttagaagg ctgccctctg ggacatttgg gtgctaactg ctcccagcca   2160
tgccagtgtg gtcctggaga aaagtgccac ccagagactg ggcctgtgt atgtccccca   2220
gggcacagtg gtgcaccttg caggattgga atccaggagc cctttactgt gatgccgacc   2280
actccagtag cgtataactc gctgggtgca gtgattggca ttgcagtgct ggggtccctt   2340
gtggtagccc tggtggcact gttcattggc tatcggcact ggcaaaaagg caaggagcac   2400
caccacctgg ctgtggctta cagcagcggg cgcctggacg gctccgagta tgtcatgcca   2460
gatgtccctc cgagctacag tcactactac tccaacccca gctaccacac cctgtcgcag   2520
tgctccccaa accccccacc ccctaacaag gttccaggcc cgctctttgc cagcctgcag   2580
aaccctgagc ggccaggtgg ggcccaaggg catgataacc acaccacccct gcctgctgac   2640
tggaagcacc gccgggagcc ccctccaggg cctctggaca gggggagcag ccgcctggac   2700
cgaagctaca gctatagcta cagcaatggc ccaggcccat tctacgataa agggctcatc   2760
tctgaagagg agctcggggc cagtgtggct tccctgagca gtgagaaccc atatgccacc   2820
atccgggacc tgcccagctt gccagggggc ccccgggaga gcagctacat ggagatgaaa   2880
ggccctccct caggatctgc ccccaggcag cctcctcagt tttgggacag ccagaggcgg   2940
cggcaacccc agccacagag agacagtggc acctacgagc agcccagccc cctgatccat   3000
gaccgagact ctgtgggctc ccagcccccct ctgcctccgg gcctaccccc cggccactat   3060
gactcaccca agaacagcca catccctgga cattatgact tgcctccagt acggcatccc   3120
ccatcacctc cacttcgacg ccaggaccgt                                      3150
```

```
<210> 16
<211> 2569
<212> DNA
<213> Mus musculus

<220>
<221> CDS
<222> (2)...(1492)
```

<400> 16

g tcg acc cac gcg tcc ggt gac cct gtt cat gga cag tgc cga tgt cag 49
Ser Thr His Ala Ser Gly Asp Pro Val His Gly Gln Cys Arg Cys Gln
1 5 10 15

gct ggt tgg atg ggc aca cgc tgc cac ctg cct tgc ccg gag ggc ttt 97
Ala Gly Trp Met Gly Thr Arg Cys His Leu Pro Cys Pro Glu Gly Phe
20 25 30

tgg gga gcc aac tgc agt aac acc tgt acc tgc aag aat ggt ggt acc 145
Trp Gly Ala Asn Cys Ser Asn Thr Cys Thr Cys Lys Asn Gly Gly Thr
35 40 45

tgt gtg tct gag aat ggc aac tgc gtg tgc gca cca ggg ttc cga ggc 193
Cys Val Ser Glu Asn Gly Asn Cys Val Cys Ala Pro Gly Phe Arg Gly
50 55 60

ccc tcc tgc cag agg ccc tgc ccg cct ggt cgc tat ggc aaa cgc tgt 241
Pro Ser Cys Gln Arg Pro Cys Pro Pro Gly Arg Tyr Gly Lys Arg Cys
65 70 75 80

gtg caa tgc aag tgt aac aac aac cat tct tcc tgc cac cca tcg gac 289
Val Gln Cys Lys Cys Asn Asn Asn His Ser Ser Cys His Pro Ser Asp
85 90 95

ggg acc tgc tcc tgc ctg gcg ggc tgg aca ggc cct gac tgc tcc gag 337
Gly Thr Cys Ser Cys Leu Ala Gly Trp Thr Gly Pro Asp Cys Ser Glu
100 105 110

gca tgt ccc cca ggc cac tgg gga ctc aaa tgc tcc caa ctc tgc cag 385
Ala Cys Pro Pro Gly His Trp Gly Leu Lys Cys Ser Gln Leu Cys Gln
115 120 125

tgt cat cat ggt ggg acc tgc cac ccc cag gat ggg agc tgt atc tgc 433
Cys His His Gly Gly Thr Cys His Pro Gln Asp Gly Ser Cys Ile Cys
130 135 140

acg cca ggc tgg act gga ccc aac tgc ttg gaa ggc tgc cca cca aga 481
Thr Pro Gly Trp Thr Gly Pro Asn Cys Leu Glu Gly Cys Pro Pro Arg
145 150 155 160

atg ttt ggt gtc aac tgc tcc cag cta tgt cag tgt gat ctc gga gag 529
Met Phe Gly Val Asn Cys Ser Gln Leu Cys Gln Cys Asp Leu Gly Glu
165 170 175

atg tgc cac cca gag act ggg gct tgt gtc tgt ccc cca gga cac agt 577
Met Cys His Pro Glu Thr Gly Ala Cys Val Cys Pro Pro Gly His Ser
180 185 190

ggt gca gac tgc aaa atg gga agc cag gag tcc ttc acc ata atg ccc 625
Gly Ala Asp Cys Lys Met Gly Ser Gln Glu Ser Phe Thr Ile Met Pro
195 200 205

acc tct ccc gtg acc cat aac tca ctg ggt gca gtg att ggc att gca 673
Thr Ser Pro Val Thr His Asn Ser Leu Gly Ala Val Ile Gly Ile Ala
210 215 220

gta ctg gga acc ctc gtg gtg gcc ctg ata gca ctg ttc att ggc tac 721

```
Val Leu Gly Thr Leu Val Val Ala Leu Ile Ala Leu Phe Ile Gly Tyr
225                 230                 235                 240

cgc cag tgg caa aag ggc aag gaa cat gag cac ttg gca gtg gct tac      769
Arg Gln Trp Gln Lys Gly Lys Glu His Glu His Leu Ala Val Ala Tyr
                245                 250                 255

agc act ggg cgg ctg gat ggc tct gat tac gtc atg cca gat gtc tct      817
Ser Thr Gly Arg Leu Asp Gly Ser Asp Tyr Val Met Pro Asp Val Ser
            260                 265                 270

ccg agc tat agt cac tac tac tcc aac ccc agc tac cac aca ctg tct      865
Pro Ser Tyr Ser His Tyr Tyr Ser Asn Pro Ser Tyr His Thr Leu Ser
        275                 280                 285

cag tgt tct cct aac ccc ccg ccc cct aac aag gtc cca ggc agt cag      913
Gln Cys Ser Pro Asn Pro Pro Pro Pro Asn Lys Val Pro Gly Ser Gln
    290                 295                 300

ctc ttt gtc agc tct cag gcc cct gag cgg cca agc aga gcc cac ggg      961
Leu Phe Val Ser Ser Gln Ala Pro Glu Arg Pro Ser Arg Ala His Gly
305                 310                 315                 320

cgt gag aac cat acc aca ctg ccc gct gac tgg aag cac cgc cgg gag     1009
Arg Glu Asn His Thr Thr Leu Pro Ala Asp Trp Lys His Arg Arg Glu
                325                 330                 335

ccc cat gac aga ggc gcc agc cac ctg gac cga agc tat agc tgt agc     1057
Pro His Asp Arg Gly Ala Ser His Leu Asp Arg Ser Tyr Ser Cys Ser
            340                 345                 350

tat agc cac agg aat ggc cca gga cca ttc tgt cat aaa ggt ccc atc     1105
Tyr Ser His Arg Asn Gly Pro Gly Pro Phe Cys His Lys Gly Pro Ile
        355                 360                 365

tct gaa gag gga cta ggg gca agc gtt atg tcc ctg agc agt gag aac     1153
Ser Glu Glu Gly Leu Gly Ala Ser Val Met Ser Leu Ser Ser Glu Asn
    370                 375                 380

ccc tat gct acc atc cga gac ctg ccc agc ctg cct ggg gaa ccc cga     1201
Pro Tyr Ala Thr Ile Arg Asp Leu Pro Ser Leu Pro Gly Glu Pro Arg
385                 390                 395                 400

gaa agt ggc tat gtg gag atg aaa gga cct cca tca gtg tcc cct ccc     1249
Glu Ser Gly Tyr Val Glu Met Lys Gly Pro Pro Ser Val Ser Pro Pro
                405                 410                 415

agg cag tct ctt cat ctc cgg gac agg cag cag cgg caa ctg cag cca     1297
Arg Gln Ser Leu His Leu Arg Asp Arg Gln Gln Arg Gln Leu Gln Pro
            420                 425                 430

cag agg gac agc ggc acc tat gag cag ccc agc ccc ttg agc cat aat     1345
Gln Arg Asp Ser Gly Thr Tyr Glu Gln Pro Ser Pro Leu Ser His Asn
        435                 440                 445

gaa gag tct ttg ggc tcc acg ccc ccg ctt cct cca ggc ctg cct cct     1393
Glu Glu Ser Leu Gly Ser Thr Pro Pro Leu Pro Pro Gly Leu Pro Pro
    450                 455                 460
```

```
ggt cac tac gac tcc ccc aag aac agc cat atc cct gga cac tat gac      1441
Gly His Tyr Asp Ser Pro Lys Asn Ser His Ile Pro Gly His Tyr Asp
465                 470                 475                 480

ttg cct cca gta cgg cat cct cca tcc cct cca tcc cgg cgc cag gac      1489
Leu Pro Pro Val Arg His Pro Pro Ser Pro Pro Ser Arg Arg Gln Asp
                485                 490                 495

cgc tgaagagccg gcatggtatg ggagcgtgcc tatgtacctt gccaggagca           1542
Arg


gggactggac cagcaggcca cgaacagaaa cacttggtga agtgaacaga gacggactgt    1602
ggccctgtgc ttccaccgag ggagacacta gttgacaaag tgtctaaccc tcttttccaa    1662
cccactgctc aagtccctgt ggacataagc tggtgggcag aatgttgttg tacaagtgtg    1722
atttagatc gatttttttt taaagtatgt gttgggtacc ttttctgtgt gtatgctcag     1782
gcaggctgtg tgtgtctcta gttggcttta gagggagtca ggtataggtt ctgccttctg    1842
cactttccat cttatctagt agtcagcttc caagcttaac tagttagagc tccaccagca    1902
gcaggcccta actacctgcc tgcccttcac ccagtaatcc tccatgtctt tgctcagagg    1962
attgctcccc gactctggtg ttgtcctcct ggtacgcctt gacggtcctg cagtctccct    2022
ttcccgtctt gcttcattct ttcccagaat gaaggctgtc tgccacccta cttcccagcc    2082
caggaattgg cacatctaag ttcagccttc ctaagttacc cgttgagtcc tgcttgccct    2142
tcacatattc cacagaacac ccaccccaca tctgcttcat agctactctc ttctccacgt    2202
acccacagaa ggcagaagtg gtaccaggca agaagatggg attgttgcat tttgttttgt    2262
ttttgagact ctgtctcact atgtagtcct ggctggcctg gaactcaaga gctctgcctg    2322
cctctgcctc ttgagtgctg ggtttaacgg ctcagggtca catgcacagc tcaagctgca    2382
ctccgatgtg ctttcccctg ttgctagatt agcgtctgcc tccccctagt ggagaggctg    2442
atcgccagct ctctgatgca ggactctggt gtttaggctc actcactatt ggtttccttg    2502
gcacagggta gtcactcaat aaatgttcct ctaaaagctg aaaaaaaaaa aaaaaaaggg    2562
cggccgc                                                              2569

      &lt;210&gt; 17
      &lt;211&gt; 497
      &lt;212&gt; PRT
      &lt;213&gt; Mus musculus

      &lt;400&gt; 17
Ser Thr His Ala Ser Gly Asp Pro Val His Gly Gln Cys Arg Cys Gln
 1               5                  10                  15
Ala Gly Trp Met Gly Thr Arg Cys His Leu Pro Cys Pro Glu Gly Phe
            20                  25                  30
Trp Gly Ala Asn Cys Ser Asn Thr Cys Thr Cys Lys Asn Gly Gly Thr
        35                  40                  45
Cys Val Ser Glu Asn Gly Asn Cys Val Cys Ala Pro Gly Phe Arg Gly
    50                  55                  60
Pro Ser Cys Gln Arg Pro Cys Pro Pro Gly Arg Tyr Gly Lys Arg Cys
65                  70                  75                  80
Val Gln Cys Lys Cys Asn Asn Asn His Ser Ser Cys His Pro Ser Asp
                85                  90                  95
Gly Thr Cys Ser Cys Leu Ala Gly Trp Thr Gly Pro Asp Cys Ser Glu
            100                 105                 110
Ala Cys Pro Pro Gly His Trp Gly Leu Lys Cys Ser Gln Leu Cys Gln
        115                 120                 125
Cys His His Gly Gly Thr Cys His Pro Gln Asp Gly Ser Cys Ile Cys
    130                 135                 140
Thr Pro Gly Trp Thr Gly Pro Asn Cys Leu Glu Gly Cys Pro Pro Arg
145                 150                 155                 160
```

```
Met Phe Gly Val Asn Cys Ser Gln Leu Cys Gln Cys Asp Leu Gly Glu
                165                 170                 175
Met Cys His Pro Glu Thr Gly Ala Cys Val Cys Pro Pro Gly His Ser
            180                 185                 190
Gly Ala Asp Cys Lys Met Gly Ser Gln Glu Ser Phe Thr Ile Met Pro
        195                 200                 205
Thr Ser Pro Val Thr His Asn Ser Leu Gly Ala Val Ile Gly Ile Ala
    210                 215                 220
Val Leu Gly Thr Leu Val Val Ala Leu Ile Ala Leu Phe Ile Gly Tyr
225                 230                 235                 240
Arg Gln Trp Gln Lys Gly Lys Glu His Glu His Leu Ala Val Ala Tyr
                245                 250                 255
Ser Thr Gly Arg Leu Asp Gly Ser Asp Tyr Val Met Pro Asp Val Ser
            260                 265                 270
Pro Ser Tyr Ser His Tyr Tyr Ser Asn Pro Ser Tyr His Thr Leu Ser
        275                 280                 285
Gln Cys Ser Pro Asn Pro Pro Pro Pro Asn Lys Val Pro Gly Ser Gln
    290                 295                 300
Leu Phe Val Ser Ser Gln Ala Pro Glu Arg Pro Ser Arg Ala His Gly
305                 310                 315                 320
Arg Glu Asn His Thr Thr Leu Pro Ala Asp Trp Lys His Arg Arg Glu
                325                 330                 335
Pro His Asp Arg Gly Ala Ser His Leu Asp Arg Ser Tyr Ser Cys Ser
            340                 345                 350
Tyr Ser His Arg Asn Gly Pro Gly Pro Phe Cys His Lys Gly Pro Ile
        355                 360                 365
Ser Glu Glu Gly Leu Gly Ala Ser Val Met Ser Leu Ser Ser Glu Asn
    370                 375                 380
Pro Tyr Ala Thr Ile Arg Asp Leu Pro Ser Leu Pro Gly Glu Pro Arg
385                 390                 395                 400
Glu Ser Gly Tyr Val Glu Met Lys Gly Pro Pro Ser Val Ser Pro Pro
                405                 410                 415
Arg Gln Ser Leu His Leu Arg Asp Arg Gln Gln Arg Gln Leu Gln Pro
            420                 425                 430
Gln Arg Asp Ser Gly Thr Tyr Glu Gln Pro Ser Pro Leu Ser His Asn
        435                 440                 445
Glu Glu Ser Leu Gly Ser Thr Pro Pro Leu Pro Pro Gly Leu Pro Pro
    450                 455                 460
Gly His Tyr Asp Ser Pro Lys Asn Ser His Ile Pro Gly His Tyr Asp
465                 470                 475                 480
Leu Pro Pro Val Arg His Pro Pro Ser Pro Pro Ser Arg Arg Gln Asp
                485                 490                 495
Arg
```

```
<210> 18
<211> 1491
<212> DNA
<213> Mus musculus

<400> 18
tcgacccacg cgtccggtga ccctgttcat ggacagtgcc gatgtcaggc tggttggatg      60
ggcacacgct gccacctgcc ttgcccggag ggcttttggg gagccaactg cagtaacacc     120
tgtacctgca agaatggtgg tacctgtgtg tctgagaatg gcaactgcgt gtgcgcacca     180
gggttccgag gcccctcctg ccagaggccc tgcccgcctg gtcgctatgg caaacgctgt     240
gtgcaatgca agtgtaacaa caaccattct tcctgccacc catcggacgg gacctgctcc     300
tgcctggcgg gctggacagg ccctgactgc tccgaggcat gtcccccagg ccactgggga     360
ctcaaatgct cccaactctg ccagtgtcat catggtggga cctgccaccc ccaggatggg     420
```

```
agctgtatct gcacgccagg ctggactgga cccaactgct tggaaggctg cccaccaaga      480
atgtttggtg tcaactgctc ccagctatgt cagtgtgatc tcggagagat gtgccaccca      540
gagactgggg cttgtgtctg tcccccagga cacagtggtg cagactgcaa aatgggaagc      600
caggagtcct tcaccataat gcccacctct cccgtgaccc ataactcact gggtgcagtg      660
attggcattg cagtactggg aaccctcgtg gtggccctga tagcactgtt cattggctac      720
cgccagtggc aaaagggcaa ggaacatgag cacttggcag tggcttacag cactgggcgg      780
ctggatggct ctgattacgt catgccagat gtctctccga gctatagtca ctactactcc      840
aaccccagct accacacact gtctcagtgt tctcctaacc ccccgccccc taacaaggtc      900
ccaggcagtc agctctttgt cagctctcag gcccctgagc ggccaagcag agcccacggg      960
cgtgagaacc ataccacact gcccgctgac tggaagcacc gccgggagcc ccatgacaga     1020
ggcgccagcc acctggaccg aagctatagc tgtagctata gccacaggaa tggcccagga     1080
ccattctgtc ataaaggtcc catctctgaa gagggactag ggcaagcgt tatgtccctg     1140
agcagtgaga accctatgc taccatccga gacctgccca gcctgcctgg ggaaccccga     1200
gaaagtggct atgtggagat gaaaggacct ccatcagtgt cccctcccag gcagtctctt     1260
catctccggg acaggcagca gcggcaactg cagccacaga gggacagcgg cacctatgag     1320
cagcccagcc ccttgagcca taatgaagag tctttgggct ccacgccccc gcttcctcca     1380
ggcctgcctc ctggtcacta cgactccccc aagaacagcc atatccctgg acactatgac     1440
ttgcctccag tacggcatcc tccatcccct ccatcccggc gccaggaccg c               1491

      <210> 19
      <211> 3567
      <212> DNA
      <213> Rauttus sp.

      <220>
      <221> CDS
      <222> (925)...(2832)

      <400> 19
gtccgaccca cgcgtccgag ccacaccctg aaggtggttg gaaggaggga aggatctagg       60
tcctgagcac tggaattccc cagaacagca tctggcttcc cagacccatg ctggccacca      120
ctgatgtgtc cttccggctg ctggctgcag tgctgttctg ttgttgggtg ccctgtggca      180
ggcttgtgca atgccactct gtcccctcct cctcctggcc ctaggcctgc gtctggctgg      240
aacactcaac tccaatgatc ccaatgtctg taccttctgg gaaagcttca ccacgaccac      300
taaggagtcc caccttcgcc ccttcagcct gcccccagcc gagtcctgcg acaggccctg      360
ggaagacccc cacacctgcg ctcagcctac ggttgtctac cggactgtgt accgtcaggt      420
ggtgaagatg gactcccgcc cacgcctgca gtgctgtggg ggttactacg agagcagtgg      480
agcctgtgtc ccactctgtg cccaggagtg tgtccacggt cgctgtgtgg ctcctaatcg      540
gtgccagtgt gcaccaggct ggcggggtga cgactgttcc agtgagtgtg ctcctggaat      600
gtggggacca cagtgtgaca ggctctgcct ctgtggcaac agcagttcct gtgatcccag      660
gagtggggtg tgtttttgcc cctctggcct gcagcccccc gactgccttc agccttgccc      720
cgatggccac tatggtcctg cctgccagtt tgattgccat tgctatgggg catcctgtga      780
cccccgggat ggagcctgct tctgcccccc agggagaaca ggacccaggg cactgatggc      840
ttcttctgcc ccagaactta tccttgccaa aatggaggtg ttcctcaggg ctctcaaggc      900
tcctgcagct gcccaccggg ctgg atg ggt gtc atc tgt tcc ctg cca tgc         951
                           Met Gly Val Ile Cys Ser Leu Pro Cys
                            1               5

cca gag ggt ttc cac gga ccc aac tgt act cag gaa tgt cgt tgc cac        999
Pro Glu Gly Phe His Gly Pro Asn Cys Thr Gln Glu Cys Arg Cys His
 10                  15                  20                  25

aat ggt ggc ctt tgt gac agg ttt act ggg cag tgc cac tgt gct cct       1047
Asn Gly Gly Leu Cys Asp Arg Phe Thr Gly Gln Cys His Cys Ala Pro
                 30                  35                  40

ggc tat atc ggg gat cgg tgc cgt gaa gag tgc cct gtg ggc cgc ttc       1095
Gly Tyr Ile Gly Asp Arg Cys Arg Glu Glu Cys Pro Val Gly Arg Phe
```

45 50 55

ggt caa gac tgt gct gag acc tgt gac tgt gct cct ggc gct cgt tgc 1143
Gly Gln Asp Cys Ala Glu Thr Cys Asp Cys Ala Pro Gly Ala Arg Cys
60 65 70

ttt cct gcc aat ggc gcg tgt ctg tgc gaa cat ggc ttc aca ggc gac 1191
Phe Pro Ala Asn Gly Ala Cys Leu Cys Glu His Gly Phe Thr Gly Asp
75 80 85

cgc tgc act gag cga ctc tgt cca gat ggc cgc tat ggt ctg agc tgc 1239
Arg Cys Thr Glu Arg Leu Cys Pro Asp Gly Arg Tyr Gly Leu Ser Cys
90 95 100 105

caa gat ccc tgc acc tgc gac cca gaa cac agt ctc agc tgc cac cca 1287
Gln Asp Pro Cys Thr Cys Asp Pro Glu His Ser Leu Ser Cys His Pro
110 115 120

atg cac ggc gag tgc tcc tgc cag cca ggt tgg gcg ggc ctc cac tgc 1335
Met His Gly Glu Cys Ser Cys Gln Pro Gly Trp Ala Gly Leu His Cys
125 130 135

aac gag agc tgc cct cag gac acg cac gga gcc ggt tgc cag gag cac 1383
Asn Glu Ser Cys Pro Gln Asp Thr His Gly Ala Gly Cys Gln Glu His
140 145 150

tgc ctc tgt ctg cac ggc ggt gtt tgc ctc gcc gac agc ggc ctc tgc 1431
Cys Leu Cys Leu His Gly Gly Val Cys Leu Ala Asp Ser Gly Leu Cys
155 160 165

cgg tgt gca cct ggc tac acg gga cct cac tgc gct aat ctt tgt cca 1479
Arg Cys Ala Pro Gly Tyr Thr Gly Pro His Cys Ala Asn Leu Cys Pro
170 175 180 185

cct aac act tat ggg atc aac tgt tcc tcc cac tgc tcc tgt gaa aat 1527
Pro Asn Thr Tyr Gly Ile Asn Cys Ser Ser His Cys Ser Cys Glu Asn
190 195 200

gcc att gcc tgc tct cct gtc gac ggc acg tgc atc tgc aag gaa ggt 1575
Ala Ile Ala Cys Ser Pro Val Asp Gly Thr Cys Ile Cys Lys Glu Gly
205 210 215

tgg cag cgt ggt aac tgc tct gtg ccc tgt ccc cct ggc acc tgg ggc 1623
Trp Gln Arg Gly Asn Cys Ser Val Pro Cys Pro Pro Gly Thr Trp Gly
220 225 230

ttc agt tgc aat gcc agt tgc cag tgt gcc cac gag gga gtc tgc agc 1671
Phe Ser Cys Asn Ala Ser Cys Gln Cys Ala His Glu Gly Val Cys Ser
235 240 245

ccc caa act gga gcc tgt act tgc acc cct ggg tgg cgt ggg gtt cac 1719
Pro Gln Thr Gly Ala Cys Thr Cys Thr Pro Gly Trp Arg Gly Val His
250 255 260 265

tgc caa ctt ccg tgc ccg aag gga cag ttt ggt gaa ggt tgt gcc agt 1767
Cys Gln Leu Pro Cys Pro Lys Gly Gln Phe Gly Glu Gly Cys Ala Ser
270 275 280

gtc tgt gac tgt gac cac tcc gat ggc tgt gac cct gtt cat gga cac 1815
Val Cys Asp Cys Asp His Ser Asp Gly Cys Asp Pro Val His Gly His
285 290 295

tgc cga tgt cag gct ggc tgg atg ggc aca cgt tgc cac ctg cct tgc 1863
Cys Arg Cys Gln Ala Gly Trp Met Gly Thr Arg Cys His Leu Pro Cys
300 305 310

cca gag ggc ttt tgg gga gcc aac tgc agc aat gcc tgt acc tgc aag 1911
Pro Glu Gly Phe Trp Gly Ala Asn Cys Ser Asn Ala Cys Thr Cys Lys
315 320 325

aat ggt ggc act tgt gta cct gag aac ggc aac tgt gtg tgc gca cca 1959
Asn Gly Gly Thr Cys Val Pro Glu Asn Gly Asn Cys Val Cys Ala Pro
330 335 340 345

ggg ttc aga ggc ccc tcc tgc cag agg ccc tgc ccg cct ggt cgc tat 2007
Gly Phe Arg Gly Pro Ser Cys Gln Arg Pro Cys Pro Pro Gly Arg Tyr
350 355 360

ggc aaa cgc tgt gtg ccc tgc aag tgc aac aac cat tct tcc tgc cac 2055
Gly Lys Arg Cys Val Pro Cys Lys Cys Asn Asn His Ser Ser Cys His
365 370 375

ccg tcg gat ggg acc tgc tcc tgc ctg gca ggc tgg aca ggc cct gac 2103
Pro Ser Asp Gly Thr Cys Ser Cys Leu Ala Gly Trp Thr Gly Pro Asp
380 385 390

tgc tct gaa tca tgt ccc cca ggc cac tgg gga ctc aaa tgc tcc caa 2151
Cys Ser Glu Ser Cys Pro Pro Gly His Trp Gly Leu Lys Cys Ser Gln
395 400 405

ccc tgc cag tgt cat cat ggt gcc acc tgc cac ccc cag gat ggg agc 2199
Pro Cys Gln Cys His His Gly Ala Thr Cys His Pro Gln Asp Gly Ser
410 415 420 425

tgt gtc tgc atc cca ggc tgg act gga ccc aac tgc tcg gaa ggc tgc 2247
Cys Val Cys Ile Pro Gly Trp Thr Gly Pro Asn Cys Ser Glu Gly Cys
430 435 440

cca tca aga atg ttt ggt gtc aac tgc tcc cag cta tgt cag tgt gat 2295
Pro Ser Arg Met Phe Gly Val Asn Cys Ser Gln Leu Cys Gln Cys Asp
445 450 455

cct gga gag atg tgc cac cca gag act ggg gct tgc gtc tgt ccc cca 2343
Pro Gly Glu Met Cys His Pro Glu Thr Gly Ala Cys Val Cys Pro Pro
460 465 470

gga cac agt ggt gcg cac tgc aaa gtg ggc agc cag gag tcc ttc acc 2391
Gly His Ser Gly Ala His Cys Lys Val Gly Ser Gln Glu Ser Phe Thr
475 480 485

ata atg ccc acc tct cct gtg atc cat aac tca ctg ggt gcc gtg att 2439
Ile Met Pro Thr Ser Pro Val Ile His Asn Ser Leu Gly Ala Val Ile
490 495 500 505

ggc att gca gtg ctg ggg acc ctt gtg gtg gcc ctg gta gca ctg ttt 2487
Gly Ile Ala Val Leu Gly Thr Leu Val Val Ala Leu Val Ala Leu Phe

```
                510                 515                 520

att ggc tac cga cac tgg caa aag ggc aag gaa cat gag cac ttg gca      2535
Ile Gly Tyr Arg His Trp Gln Lys Gly Lys Glu His Glu His Leu Ala
            525                 530                 535

gtg gct tac agc act ggg cga ctg gat ggc tcc gat tac gtc atg cca      2583
Val Ala Tyr Ser Thr Gly Arg Leu Asp Gly Ser Asp Tyr Val Met Pro
        540                 545                 550

gat gtc tct ccg agc tac agt cac tac tat tcc aac cct agc tac cac      2631
Asp Val Ser Pro Ser Tyr Ser His Tyr Tyr Ser Asn Pro Ser Tyr His
    555                 560                 565

aca ctg tct cag tgt tct cct aac cct cca ccc cct aac aag att cca      2679
Thr Leu Ser Gln Cys Ser Pro Asn Pro Pro Pro Pro Asn Lys Ile Pro
570                 575                 580                 585

ggc agt cag ctg ttt gtc agc tcc cag gca tct gag cgg cca aac aga      2727
Gly Ser Gln Leu Phe Val Ser Ser Gln Ala Ser Glu Arg Pro Asn Arg
                590                 595                 600

aac cat ggg cga gat aac cac gcc aca ctg ccc gct gac tgg aag cac      2775
Asn His Gly Arg Asp Asn His Ala Thr Leu Pro Ala Asp Trp Lys His
            605                 610                 615

cga cgg gag tcc cat gac aga gct ttc ctc agg cac cag cca cct gga      2823
Arg Arg Glu Ser His Asp Arg Ala Phe Leu Arg His Gln Pro Pro Gly
        620                 625                 630

ccg aag gta tagctgtagc tatggccaca ggaatggccc ggggccattc              2872
Pro Lys Val
    635

tgtcataaag gtcccatctc tgaagaagga ctaggggcaa gcgttatgtc cctgagcagt   2932
gagaacccct atgcgaccat ccgagacctg cccggcctgc ctggggaacc ccgagaaagc   2992
agctatgtgg agatgaaagg ccctccatca gtgtctcccc ccaggcagcc tcttcatctc   3052
cgggacaggc agcagcagca actgcagtct cagagagaca gcggcaccta tgagcagccc   3112
actcccttga gccgtaatga agagtctgtg ggctccatgc cccctcttcc tccgggcctg   3172
ccacccggcc actatgactc gcccaaaaac agccacatcc ctggacacta tgacttgcct   3232
ccagtacggc atcctccatc acctccatcc cggcgccagg accgctgagg agccagcatg   3292
gtatgggaga gtgcctgtga accctgccag gagcagggcc tggaccagca ggccatgaat   3352
agacatactt ggtgaagtga acggagactg aggatggctc tgcttccacc gagggagaca   3412
ctagttggca aagtgtctaa cctccctttt ccagcccatt gctcaagtcc cccaggctgt   3472
ggacatgagc tggtgggcag aatgttgttg ttgaagtctg attttagatt gatttttttaa  3532
aaaaaaaaaa aaaaaaaaaa aaaaagggcg gccgc                               3567

      <210> 20
      <211> 636
      <212> PRT
      <213> Rauttus sp.

      <400> 20
Met Gly Val Ile Cys Ser Leu Pro Cys Pro Glu Gly Phe His Gly Pro
 1               5                  10                  15
Asn Cys Thr Gln Glu Cys Arg Cys His Asn Gly Gly Leu Cys Asp Arg
            20                  25                  30
Phe Thr Gly Gln Cys His Cys Ala Pro Gly Tyr Ile Gly Asp Arg Cys
```

35 40 45
Arg Glu Glu Cys Pro Val Gly Arg Phe Gly Gln Asp Cys Ala Glu Thr
50 55 60
Cys Asp Cys Ala Pro Gly Ala Arg Cys Phe Pro Ala Asn Gly Ala Cys
65 70 75 80
Leu Cys Glu His Gly Phe Thr Gly Asp Arg Cys Thr Glu Arg Leu Cys
85 90 95
Pro Asp Gly Arg Tyr Gly Leu Ser Cys Gln Asp Pro Cys Thr Cys Asp
100 105 110
Pro Glu His Ser Leu Ser Cys His Pro Met His Gly Glu Cys Ser Cys
115 120 125
Gln Pro Gly Trp Ala Gly Leu His Cys Asn Glu Ser Cys Pro Gln Asp
130 135 140
Thr His Gly Ala Gly Cys Gln Glu His Cys Leu Cys Leu His Gly Gly
145 150 155 160
Val Cys Leu Ala Asp Ser Gly Leu Cys Arg Cys Ala Pro Gly Tyr Thr
165 170 175
Gly Pro His Cys Ala Asn Leu Cys Pro Pro Asn Thr Tyr Gly Ile Asn
180 185 190
Cys Ser Ser His Cys Ser Cys Glu Asn Ala Ile Ala Cys Ser Pro Val
195 200 205
Asp Gly Thr Cys Ile Cys Lys Glu Gly Trp Gln Arg Gly Asn Cys Ser
210 215 220
Val Pro Cys Pro Pro Gly Thr Trp Gly Phe Ser Cys Asn Ala Ser Cys
225 230 235 240
Gln Cys Ala His Glu Gly Val Cys Ser Pro Gln Thr Gly Ala Cys Thr
245 250 255
Cys Thr Pro Gly Trp Arg Gly Val His Cys Gln Leu Pro Cys Pro Lys
260 265 270
Gly Gln Phe Gly Glu Gly Cys Ala Ser Val Cys Asp Cys Asp His Ser
275 280 285
Asp Gly Cys Asp Pro Val His Gly His Cys Arg Cys Gln Ala Gly Trp
290 295 300
Met Gly Thr Arg Cys His Leu Pro Cys Pro Glu Gly Phe Trp Gly Ala
305 310 315 320
Asn Cys Ser Asn Ala Cys Thr Cys Lys Asn Gly Gly Thr Cys Val Pro
325 330 335
Glu Asn Gly Asn Cys Val Cys Ala Pro Gly Phe Arg Gly Pro Ser Cys
340 345 350
Gln Arg Pro Cys Pro Pro Gly Arg Tyr Gly Lys Arg Cys Val Pro Cys
355 360 365
Lys Cys Asn Asn His Ser Ser Cys His Pro Ser Asp Gly Thr Cys Ser
370 375 380
Cys Leu Ala Gly Trp Thr Gly Pro Asp Cys Ser Glu Ser Cys Pro Pro
385 390 395 400
Gly His Trp Gly Leu Lys Cys Ser Gln Pro Cys Gln Cys His His Gly
405 410 415
Ala Thr Cys His Pro Gln Asp Gly Ser Cys Val Cys Ile Pro Gly Trp
420 425 430
Thr Gly Pro Asn Cys Ser Glu Gly Cys Pro Ser Arg Met Phe Gly Val
435 440 445
Asn Cys Ser Gln Leu Cys Gln Cys Asp Pro Gly Glu Met Cys His Pro
450 455 460
Glu Thr Gly Ala Cys Val Cys Pro Pro Gly His Ser Gly Ala His Cys
465 470 475 480
Lys Val Gly Ser Gln Glu Ser Phe Thr Ile Met Pro Thr Ser Pro Val
485 490 495
Ile His Asn Ser Leu Gly Ala Val Ile Gly Ile Ala Val Leu Gly Thr

```
            500                 505                 510
Leu Val Val Ala Leu Val Ala Leu Phe Ile Gly Tyr Arg His Trp Gln
        515                 520                 525
Lys Gly Lys Glu His Glu His Leu Ala Val Ala Tyr Ser Thr Gly Arg
    530                 535                 540
Leu Asp Gly Ser Asp Tyr Val Met Pro Asp Val Ser Pro Ser Tyr Ser
545                 550                 555                 560
His Tyr Tyr Ser Asn Pro Ser Tyr His Thr Leu Ser Gln Cys Ser Pro
                565                 570                 575
Asn Pro Pro Pro Pro Asn Lys Ile Pro Gly Ser Gln Leu Phe Val Ser
            580                 585                 590
Ser Gln Ala Ser Glu Arg Pro Asn Arg Asn His Gly Arg Asp Asn His
        595                 600                 605
Ala Thr Leu Pro Ala Asp Trp Lys His Arg Arg Glu Ser His Asp Arg
    610                 615                 620
Ala Phe Leu Arg His Gln Pro Pro Gly Pro Lys Val
625                 630                 635

      <210> 21
      <211> 1908
      <212> DNA
      <213> Rattus sp.

      <400> 21
atgggtgtca tctgttccct gccatgccca gagggtttcc acggacccaa ctgtactcag      60
gaatgtcgtt gccacaatgg tggcctttgt gacaggttta ctgggcagtg ccactgtgct     120
cctggctata tcggggatcg gtgccgtgaa gagtgccctg tgggccgctt cggtcaagac     180
tgtgctgaga cctgtgactg tgctcctggc gctcgttgct ttcctgccaa tggcgcgtgt     240
ctgtgcgaac atggcttcac aggcgaccgc tgcactgagc gactctgtcc agatggccgc     300
tatggtctga gctgccaaga tccctgcacc tgcgacccag aacacagtct cagctgccac     360
ccaatgcacg gcgagtgctc ctgccagcca ggttgggcgg gcctccactg caacgagagc     420
tgccctcagg acacgcacgg agccggttgc caggagcact gcctctgtct gcacggcggt     480
gtttgcctcg ccgacagcgg cctctgccgg tgtgcacctg gctacacggg acctcactgc     540
gctaatcttt gtccacctaa cacttatggg atcaactgtt cctcccactg ctcctgtgaa     600
aatgccattg cctgctctcc tgtcgacggc acgtgcatct gcaaggaagg ttggcagcgt     660
ggtaactgct ctgtgccctg tccccctggc acctggggct tcagttgcaa tgccagttgc     720
cagtgtgccc acgagggagt ctgcagcccc caaactggag cctgtacttg cacccctggg     780
tggcgtgggg ttcactgcca acttccgtgc ccgaagggac agtttggtga aggttgtgcc     840
agtgtctgtg actgtgacca ctccgatggc tgtgaccctg ttcatggaca ctgccgatgt     900
caggctggct ggatgggcac acgttgccac ctgccttgcc cagagggctt ttggggagcc     960
aactgcagca atgcctgtac ctgcaagaat ggtggcactt gtgtacctga gaacggcaac    1020
tgtgtgtgcg caccagggtt cagaggcccc tcctgccaga ggccctgccc gcctggtcgc    1080
tatggcaaac gctgtgtgcc ctgcaagtgc aacaaccatt cttcctgcca cccgtcggat    1140
gggacctgct cctgcctggc aggctggaca ggccctgact gctctgaatc atgtccccca    1200
ggccactggg gactcaaatg ctcccaaccc tgccagtgtc atcatggtgc cacctgccac    1260
ccccaggatg ggagctgtgt ctgcatccca ggctggactg gacccaactg ctcggaaggc    1320
tgcccatcaa gaatgtttgg tgtcaactgc tcccagctat gtcagtgtga tcctggagag    1380
atgtgccacc cagagactgg ggcttgcgtc tgtcccccag gacacagtgg tgcgcactgc    1440
aaagtgggca gccaggagtc cttcaccata atgcccacct ctcctgtgat ccataactca    1500
ctgggtgccg tgattggcat tgcagtgctg gggacccttg tggtggccct ggtagcactg    1560
tttattggct accgacactg gcaaaagggc aaggaacatg agcacttggc agtggcttac    1620
agcactgggc gactggatgg ctccgattac gtcatgccag atgtctctcc gagctacagt    1680
cactactatt ccaaccctag ctaccacaca ctgtctcagt gttctcctaa ccctccaccc    1740
cctaacaaga ttccaggcag tcagctgttt gtcagctccc aggcatctga gcggccaaac    1800
agaaaccatg ggcgagataa ccacgccaca ctgcccgctg actggaagca ccgacgggag    1860
tcccatgaca gagctttcct caggcaccag ccacctggac cgaaggta                 1908

      <210> 22
```

<211> 1497
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (217)...(684)

<400> 22
gtcgacccac gcgtccggct cccagcccac ccccaaacag acacagcgta gcccgggcca 60
gctcttaagg agttcaggag tgagaagagg ccctcagaga tctgacagcc taggagtgcg 120
tggacaccac ctcagcccac tgagcaggag tcacagcacg aagaccaagc gcaaagcgac 180
ccctgccctc catcctgact gctcctccta agagag atg gca ccg gcc aga gca 234
Met Ala Pro Ala Arg Ala
1 5

gga ttc tgc ccc ctt ctg ctg ctt ctg ctg ctg ggg ctg tgg gtg gca 282
Gly Phe Cys Pro Leu Leu Leu Leu Leu Leu Leu Gly Leu Trp Val Ala
10 15 20

gag atc cca gtc agt gcc aag ccc aag ggc atg acc tca tca cag tgg 330
Glu Ile Pro Val Ser Ala Lys Pro Lys Gly Met Thr Ser Ser Gln Trp
25 30 35

ttt aaa att cag cac atg cag ccc agc cct caa gca tgc aac tca gcc 378
Phe Lys Ile Gln His Met Gln Pro Ser Pro Gln Ala Cys Asn Ser Ala
40 45 50

atg aaa aac att aac aag cac aca aaa cgg tgc aaa gac ctc aac acc 426
Met Lys Asn Ile Asn Lys His Thr Lys Arg Cys Lys Asp Leu Asn Thr
55 60 65 70

ttc ctg cac gag cct ttc tcc agt gtg gcc gcc acc tgc cag acc ccc 474
Phe Leu His Glu Pro Phe Ser Ser Val Ala Ala Thr Cys Gln Thr Pro
75 80 85

aaa ata gcc tgc aag aat ggc gat aaa aac tgc cac cag agc cac ggg 522
Lys Ile Ala Cys Lys Asn Gly Asp Lys Asn Cys His Gln Ser His Gly
90 95 100

ccc gtg tcc ctg acc atg tgt aag ctc acc tca ggg aag tat ccg aac 570
Pro Val Ser Leu Thr Met Cys Lys Leu Thr Ser Gly Lys Tyr Pro Asn
105 110 115

tgc agg tac aaa gag aag cga cag aac aag tct tac gta gtg gcc tgt 618
Cys Arg Tyr Lys Glu Lys Arg Gln Asn Lys Ser Tyr Val Val Ala Cys
120 125 130

aag cct ccc cag aaa aag gac tct cag caa ttc cac ctg gtt cct gta 666
Lys Pro Pro Gln Lys Lys Asp Ser Gln Gln Phe His Leu Val Pro Val
135 140 145 150

cac ttg gac aga gtc ctt taggtttcca gactggcttg ctctttggct 714
His Leu Asp Arg Val Leu
155

gaccttcaat tccctctcca ggactccgca ccactcccct acacccagag cattctcttc 774
ccctcatctc ttggggctgt tcctggttca gcctctgctg ggaggctgaa gctgacactc 834

tggtgagctg agctctagag ggatggcttt tcatcttttt gttgctgttt tcccagatgc 894
ttatccccaa gaaacagcaa gctcaggtct gtgggttccc tggtctatgc cattgcacat 954
gtctcccctg cccccctggca ttagggcagc atgacaagga gaggaaataa atggaaaggg 1014
ggcatatggg atttgtggac acagctgttt ctgttcctga actagaagtc ttccccagct 1074
ctgacgtggc agtgaggtga cctgaaggaa agaaaaatat aaataaatac cacttcatat 1134
ttgtatagaa tcctctaatc ccttgtgaca tagacttgac agggattgta tgccttcttt 1194
atggatgagg aaattaaggt tttagaaagc ttaatgaatt aaagagcttg tctaattagt 1254
tagtagcaga acctggactt gaacctaggt ctccttgctc taaatacagt gtaccttcta 1314
ctctaccagt tgcgcaagaa agaagtcact gttacagagg caagcggtga actaggtaag 1374
agttcactca tgaagaaacg agtgctctga agagccagtt accctgtgtt ggctgcaata 1434
aaggtcatta cctctctagc caaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1494
aaa 1497

<210> 23
<211> 156
<212> PRT
<213> Homo sapiens

<400> 23
Met Ala Pro Ala Arg Ala Gly Phe Cys Pro Leu Leu Leu Leu Leu Leu
1 5 10 15
Leu Gly Leu Trp Val Ala Glu Ile Pro Val Ser Ala Lys Pro Lys Gly
20 25 30
Met Thr Ser Ser Gln Trp Phe Lys Ile Gln His Met Gln Pro Ser Pro
35 40 45
Gln Ala Cys Asn Ser Ala Met Lys Asn Ile Asn Lys His Thr Lys Arg
50 55 60
Cys Lys Asp Leu Asn Thr Phe Leu His Glu Pro Phe Ser Ser Val Ala
65 70 75 80
Ala Thr Cys Gln Thr Pro Lys Ile Ala Cys Lys Asn Gly Asp Lys Asn
85 90 95
Cys His Gln Ser His Gly Pro Val Ser Leu Thr Met Cys Lys Leu Thr
100 105 110
Ser Gly Lys Tyr Pro Asn Cys Arg Tyr Lys Glu Lys Arg Gln Asn Lys
115 120 125
Ser Tyr Val Val Ala Cys Lys Pro Pro Gln Lys Lys Asp Ser Gln Gln
130 135 140
Phe His Leu Val Pro Val His Leu Asp Arg Val Leu
145 150 155

<210> 24
<211> 468
<212> DNA
<213> Homo sapiens

<400> 24
atggcaccgg ccagagcagg attctgcccc cttctgctgc ttctgctgct ggggctgtgg 60
gtggcagaga tcccagtcag tgccaagccc aagggcatga cctcatcaca gtggtttaaa 120
attcagcaca tgcagcccag ccctcaagca tgcaactcag ccatgaaaaa cattaacaag 180
cacacaaaac ggtgcaaaga cctcaacacc ttcctgcacg agcctttctc cagtgtggcc 240
gccacctgcc agacccccaa aatagcctgc aagaatggcg ataaaaactg ccaccagagc 300
cacgggcccg tgtccctgac catgtgtaag ctcacctcag ggaagtatcc gaactgcagg 360
tacaaagaga agcgacagaa caagtcttac gtagtggcct gtaagcctcc ccagaaaaag 420
gactctcagc aattccacct ggttcctgta cacttggaca gagtcctt 468

<210> 25
<211> 1788
<212> DNA

<213> Homo sapiens

<220>
<221> CDS
<222> (62)...(976)

<400> 25
```
gtcgacccac ggcgtccggc caggctccac tgaggggaac ggggacctgt ctgaagagaa      60
g atg ccc ctg ctg aca ctc tac ctg ctc ctc ttc tgg ctc tca ggc tac     109
  Met Pro Leu Leu Thr Leu Tyr Leu Leu Leu Phe Trp Leu Ser Gly Tyr
   1               5                  10                  15

tcc att gcc act caa atc acc ggt cca aca aca gtg aat ggc ttg gag       157
Ser Ile Ala Thr Gln Ile Thr Gly Pro Thr Thr Val Asn Gly Leu Glu
            20                  25                  30

cgg ggc tcc ttg acc gtg cag tgt gtt tac aga tca ggc tgg gag acc       205
Arg Gly Ser Leu Thr Val Gln Cys Val Tyr Arg Ser Gly Trp Glu Thr
        35                  40                  45

tac ttg aag tgg tgg tgt cga gga gct att tgg cgt gac tgc aag atc       253
Tyr Leu Lys Trp Trp Cys Arg Gly Ala Ile Trp Arg Asp Cys Lys Ile
    50                  55                  60

ctt gtt aaa acc agt ggg tca gag cag gag gtg aag agg gac cgg gtg       301
Leu Val Lys Thr Ser Gly Ser Glu Gln Glu Val Lys Arg Asp Arg Val
 65                  70                  75                  80

tcc atc aag gac aat cag aaa aac cgc acg ttc act gtg acc atg gag       349
Ser Ile Lys Asp Asn Gln Lys Asn Arg Thr Phe Thr Val Thr Met Glu
                 85                  90                  95

gat ctc atg aaa act gat gct gac act tac tgg tgt gga att gag aaa       397
Asp Leu Met Lys Thr Asp Ala Asp Thr Tyr Trp Cys Gly Ile Glu Lys
            100                 105                 110

act gga aat gac ctt ggg gtc aca gtt caa gtg acc att gac cca gcg       445
Thr Gly Asn Asp Leu Gly Val Thr Val Gln Val Thr Ile Asp Pro Ala
        115                 120                 125

tcg act cct gcc ccc acc acg cct act tcc act acg ttt aca gca cca       493
Ser Thr Pro Ala Pro Thr Thr Pro Thr Ser Thr Thr Phe Thr Ala Pro
    130                 135                 140

gtc acc caa gaa gaa act agc agc tcc cca act ctg acc ggc cac cac       541
Val Thr Gln Glu Glu Thr Ser Ser Ser Pro Thr Leu Thr Gly His His
145                 150                 155                 160

ttg gac aac agg cac aag ctc ctg aag ctc agt gtc ctc ctg ccc ctc       589
Leu Asp Asn Arg His Lys Leu Leu Lys Leu Ser Val Leu Leu Pro Leu
                165                 170                 175

atc ttc acc ata ttg ctg ctg ctt ttg gtg gcc gcc tca ctc ttg gct       637
Ile Phe Thr Ile Leu Leu Leu Leu Leu Val Ala Ala Ser Leu Leu Ala
            180                 185                 190

tgg agg atg atg aag tac cag cag aaa gca gcc ggg atg tcc cca gag       685
Trp Arg Met Met Lys Tyr Gln Gln Lys Ala Ala Gly Met Ser Pro Glu
```

195 200 205

cag gta ctg cag ccc ctg gag ggc gac ctc tgc tat gca gac ctg acc 733
Gln Val Leu Gln Pro Leu Glu Gly Asp Leu Cys Tyr Ala Asp Leu Thr
210 215 220

ctg cag ctg gcc gga acc tcc ccg cga aag gct acc acg aag ctt tcc 781
Leu Gln Leu Ala Gly Thr Ser Pro Arg Lys Ala Thr Thr Lys Leu Ser
225 230 235 240

tct gcc cag gtt gac cag gtg gaa gtg gaa tat gtc acc atg gct tcc 829
Ser Ala Gln Val Asp Gln Val Glu Val Glu Tyr Val Thr Met Ala Ser
245 250 255

ttg ccg aag gag gac att tcc tat gca tct ctg acc ttg ggt gct gag 877
Leu Pro Lys Glu Asp Ile Ser Tyr Ala Ser Leu Thr Leu Gly Ala Glu
260 265 270

gat cag gaa ccg acc tac tgc aac atg ggc cac ctc agt agc cac ctc 925
Asp Gln Glu Pro Thr Tyr Cys Asn Met Gly His Leu Ser Ser His Leu
275 280 285

ccc ggc agg ggc cct gag gag ccc acg gaa tac agc acc atc agc agg 973
Pro Gly Arg Gly Pro Glu Glu Pro Thr Glu Tyr Ser Thr Ile Ser Arg
290 295 300

cct tagcctgcac tccaggctcc ttcttggacc ccaggctgtg agcacactcc 1026
Pro
305

tgcctcatcg accgtctgcc ccctgctccc ctcatcagga ccaacccggg gactggtgcc 1086
tctgcctgat cagccagcat tgcccctagc tctgggttgg gcttggggcc aagtctcagg 1146
gggcttctag gagttggggt ttctaaacg tcccctcctc tcctacatag ttgaggaggg 1206
ggctagggat atgctctggg gctttcatgg gaatgatgaa gatgataatg agaaaaatgt 1266
tatcatttatt atcatgaagt accattatca taatacaatg aacctttatt tattgcctac 1326
cacatgttat gggctgaata atggccccca aagatatctg tgtcctaatc ctcagaactt 1386
gtgactgtta ccttctgtgg cagaaaggga cagtgcagat gtatgtaagt taaggacttt 1446
gagatagaga ggttattctt gctgattcag gtgggcccaa aatatcacca caagggtcct 1506
cataagaaag aggccagaag gtcaaagagg tagagacaaa gtgatgatgg aagtggacgt 1566
gggtgtgacg tgagcagggg ccatgaatgc cgcagccttc agatgccaga aagggaaagg 1626
aatggattcc cctgcctgga gcctccaaaa gaaaccagcc ctgcccacgc cttgacttga 1686
gcccattgaa actgatcttg agctcctggc ctccagaatt gcaggagaat aaatttgtgt 1746
tgtttttaaa aaaaaaaaaa aaaaaaaagg gcggccgcta ga 1788

<210> 26
<211> 305
<212> PRT
<213> Homo sapiens

<400> 26
Met Pro Leu Leu Thr Leu Tyr Leu Leu Leu Phe Trp Leu Ser Gly Tyr
1 5 10 15
Ser Ile Ala Thr Gln Ile Thr Gly Pro Thr Thr Val Asn Gly Leu Glu
20 25 30
Arg Gly Ser Leu Thr Val Gln Cys Val Tyr Arg Ser Gly Trp Glu Thr
35 40 45
Tyr Leu Lys Trp Trp Cys Arg Gly Ala Ile Trp Arg Asp Cys Lys Ile
50 55 60

```
Leu Val Lys Thr Ser Gly Ser Glu Gln Glu Val Lys Arg Asp Arg Val
65                  70                  75                  80
Ser Ile Lys Asp Asn Gln Lys Asn Arg Thr Phe Thr Val Thr Met Glu
                85                  90                  95
Asp Leu Met Lys Thr Asp Ala Asp Thr Tyr Trp Cys Gly Ile Glu Lys
            100                 105                 110
Thr Gly Asn Asp Leu Gly Val Thr Val Gln Val Thr Ile Asp Pro Ala
        115                 120                 125
Ser Thr Pro Ala Pro Thr Thr Pro Thr Ser Thr Thr Phe Thr Ala Pro
    130                 135                 140
Val Thr Gln Glu Glu Thr Ser Ser Ser Pro Thr Leu Thr Gly His His
145                 150                 155                 160
Leu Asp Asn Arg His Lys Leu Leu Lys Leu Ser Val Leu Leu Pro Leu
                165                 170                 175
Ile Phe Thr Ile Leu Leu Leu Leu Leu Val Ala Ala Ser Leu Leu Ala
            180                 185                 190
Trp Arg Met Met Lys Tyr Gln Gln Lys Ala Ala Gly Met Ser Pro Glu
        195                 200                 205
Gln Val Leu Gln Pro Leu Glu Gly Asp Leu Cys Tyr Ala Asp Leu Thr
    210                 215                 220
Leu Gln Leu Ala Gly Thr Ser Pro Arg Lys Ala Thr Thr Lys Leu Ser
225                 230                 235                 240
Ser Ala Gln Val Asp Gln Val Glu Val Glu Tyr Val Thr Met Ala Ser
                245                 250                 255
Leu Pro Lys Glu Asp Ile Ser Tyr Ala Ser Leu Thr Leu Gly Ala Glu
            260                 265                 270
Asp Gln Glu Pro Thr Tyr Cys Asn Met Gly His Leu Ser Ser His Leu
        275                 280                 285
Pro Gly Arg Gly Pro Glu Glu Pro Thr Glu Tyr Ser Thr Ile Ser Arg
    290                 295                 300
Pro
305
```

<210> 27
<211> 915
<212> DNA
<213> Homo sapiens

<400> 27

```
atgcccctgc tgacactcta cctgctcctc ttctggctct caggctactc cattgccact     60
caaatcaccg gtccaacaac agtgaatggc ttggagcggg gctccttgac cgtgcagtgt    120
gtttacagat caggctggga gacctacttg aagtggtggt gtcgaggagc tatttggcgt    180
gactgcaaga tccttgttaa aaccagtggg tcagagcagg aggtgaagag ggaccgggtg    240
tccatcaagg acaatcagaa aaaccgcacg ttcactgtga ccatggagga tctcatgaaa    300
actgatgctg acacttactg gtgtggaatt gagaaaactg gaaatgacct tggggtcaca    360
gttcaagtga ccattgaccc agcgtcgact cctgccccca ccacgcctac ttccactacg    420
tttacagcac cagtcaccca agaagaaact agcagctccc caactctgac cggccaccac    480
ttggacaaca ggcacaagct cctgaagctc agtgtcctcc tgcccctcat cttcaccata    540
ttgctgctgc ttttggtggc cgcctcactc ttggcttgga ggatgatgaa gtaccagcag    600
aaagcagccg ggatgtcccc agagcaggta ctgcagcccc tggagggcga cctctgctat    660
gcagacctga ccctgcagct ggccggaacc tccccgcgaa aggctaccac gaagctttcc    720
tctgcccagg ttgaccaggt ggaagtggaa tatgtcacca tggcttcctt gccgaaggag    780
gacatttcct atgcatctct gaccttgggt gctgaggatc aggaaccgac ctactgcaac    840
atgggccacc tcagtagcca cctccccggc aggggccctg aggagcccac ggaatacagc    900
accatcagca ggcct                                                     915
```

<210> 28
<211> 3258

```
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (42)...(1625)

<400> 28
cacgcgtccg gccagttctt ggaggagact ctgcacaggg c atg gat cac tgt ggt      56
                                              Met Asp His Cys Gly
                                               1               5

gcc ctt ttc ctg tgc ctg tgc ctt ctg act ttg cag aat gca aca aca       104
Ala Leu Phe Leu Cys Leu Cys Leu Leu Thr Leu Gln Asn Ala Thr Thr
                 10                  15                  20

gag aca tgg gaa gaa ctc ctg agc tac atg gag aat atg cag gtg tcc       152
Glu Thr Trp Glu Glu Leu Leu Ser Tyr Met Glu Asn Met Gln Val Ser
             25                  30                  35

agg ggc cgg agc tca gtt ttt tcc tct cgt caa ctc cac cag ctg gag       200
Arg Gly Arg Ser Ser Val Phe Ser Ser Arg Gln Leu His Gln Leu Glu
         40                  45                  50

cag atg cta ctg aac acc agc ttc cca ggc tac aac ctg acc ttg cag       248
Gln Met Leu Leu Asn Thr Ser Phe Pro Gly Tyr Asn Leu Thr Leu Gln
     55                  60                  65

aca ccc acc atc cag tct ctg gcc ttc aag ctg agc tgt gac ttc tct       296
Thr Pro Thr Ile Gln Ser Leu Ala Phe Lys Leu Ser Cys Asp Phe Ser
 70                  75                  80                  85

ggc ctc tcg ctg acc agt gcc act ctg aag cgg gtg ccc cag gca gga       344
Gly Leu Ser Leu Thr Ser Ala Thr Leu Lys Arg Val Pro Gln Ala Gly
                 90                  95                 100

ggt cag cat gcc cgg ggt cag cac gcc atg cag ttc ccc gcc gag ctg       392
Gly Gln His Ala Arg Gly Gln His Ala Met Gln Phe Pro Ala Glu Leu
            105                 110                 115

acc cgg gac gcc tgc aag acc cgc ccc agg gag ctg cgg ctc atc tgt       440
Thr Arg Asp Ala Cys Lys Thr Arg Pro Arg Glu Leu Arg Leu Ile Cys
        120                 125                 130

atc tac ttc tcc aac acc cac ttt ttc aag gat gaa aac aac tca tct       488
Ile Tyr Phe Ser Asn Thr His Phe Phe Lys Asp Glu Asn Asn Ser Ser
    135                 140                 145

ctg ctg aat aac tac gtc ctg ggg gcc cag ctg agt cat ggg cac gtg       536
Leu Leu Asn Asn Tyr Val Leu Gly Ala Gln Leu Ser His Gly His Val
150                 155                 160                 165

aac aac ctc agg gat cct gtg aac atc agc ttc tgg cac aac caa agc       584
Asn Asn Leu Arg Asp Pro Val Asn Ile Ser Phe Trp His Asn Gln Ser
                170                 175                 180

ctg gaa ggc tac acc ctg acc tgt gtc ttc tgg aag gag gga gcc agg       632
Leu Glu Gly Tyr Thr Leu Thr Cys Val Phe Trp Lys Glu Gly Ala Arg
```

```
            185                 190                 195

aaa cag ccc tgg ggg ggc tgg agc cct gag ggc tgt cgt aca gag cag      680
Lys Gln Pro Trp Gly Gly Trp Ser Pro Glu Gly Cys Arg Thr Glu Gln
        200                 205                 210

ccc tcc cac tct cag gtg ctc tgc cgc tgc aac cac ctc acc tac ttt      728
Pro Ser His Ser Gln Val Leu Cys Arg Cys Asn His Leu Thr Tyr Phe
    215                 220                 225

gct gtt ctc atg caa ctc tcc cca gcc ctg gtc cct gca gag ttg ctg      776
Ala Val Leu Met Gln Leu Ser Pro Ala Leu Val Pro Ala Glu Leu Leu
230                 235                 240                 245

gca cct ctt acg tac atc tcc ctc gtg ggc tgc agc atc tcc atc gtg      824
Ala Pro Leu Thr Tyr Ile Ser Leu Val Gly Cys Ser Ile Ser Ile Val
                250                 255                 260

gcc tcg ctg atc aca gtc ctg ctg cac ttc cat ttc agg aag cag agt      872
Ala Ser Leu Ile Thr Val Leu Leu His Phe His Phe Arg Lys Gln Ser
            265                 270                 275

gac tcc tta aca cgc atc cac atg aac ctg cat gcc tcc gtg ctg ctc      920
Asp Ser Leu Thr Arg Ile His Met Asn Leu His Ala Ser Val Leu Leu
        280                 285                 290

ctg aac atc gcc ttc ctg ctg agc ccc gca ttc gca atg tct cct gtg      968
Leu Asn Ile Ala Phe Leu Leu Ser Pro Ala Phe Ala Met Ser Pro Val
    295                 300                 305

ccc ggg tca gca tgc acg gct ctg gcc gct gcc ctg cac tac gcg ctg     1016
Pro Gly Ser Ala Cys Thr Ala Leu Ala Ala Ala Leu His Tyr Ala Leu
310                 315                 320                 325

ctc agc tgc ctc acc tgg atg gcc atc gag ggc ttc aac ctc tac ctc     1064
Leu Ser Cys Leu Thr Trp Met Ala Ile Glu Gly Phe Asn Leu Tyr Leu
                330                 335                 340

ctc ctc ggg cgt gtc tac aac atc tac atc cgc aga tat gtg ttc aag     1112
Leu Leu Gly Arg Val Tyr Asn Ile Tyr Ile Arg Arg Tyr Val Phe Lys
            345                 350                 355

ctt ggt gtg cta ggc tgg ggg gcc cca gcc ctc ctg gtg ctg ctt tcc     1160
Leu Gly Val Leu Gly Trp Gly Ala Pro Ala Leu Leu Val Leu Leu Ser
        360                 365                 370

ctc tct gtc aag agc tcg gta tac gga ccc tgc aca atc ccc gtc ttc     1208
Leu Ser Val Lys Ser Ser Val Tyr Gly Pro Cys Thr Ile Pro Val Phe
    375                 380                 385

gac agc tgg gag aat ggc aca ggc ttc cag aac atg tcc ata tgc tgg     1256
Asp Ser Trp Glu Asn Gly Thr Gly Phe Gln Asn Met Ser Ile Cys Trp
390                 395                 400                 405

gtg cgg agc ccc gtg gtg cac agt gtc ctg gtc atg ggc tac ggc ggc     1304
Val Arg Ser Pro Val Val His Ser Val Leu Val Met Gly Tyr Gly Gly
                410                 415                 420
```

```
ctc acg tcc ctc ttc aac ctg gtg gtg ctg gcc tgg gcg ctg tgg acc      1352
Leu Thr Ser Leu Phe Asn Leu Val Val Leu Ala Trp Ala Leu Trp Thr
            425                 430                 435

ctg cgc agg ctg cgg gag cgg gcg gat gca cca agt gtc agg gcc tgc      1400
Leu Arg Arg Leu Arg Glu Arg Ala Asp Ala Pro Ser Val Arg Ala Cys
        440                 445                 450

cat gac act gtc act gtg ctg ggc ctc acc gtg ctg ctg gga acc acc      1448
His Asp Thr Val Thr Val Leu Gly Leu Thr Val Leu Leu Gly Thr Thr
    455                 460                 465

tgg gcc ttg gcc ttc ttt tct ttt ggc gtc ttc ctg ctg ccc cag ctg      1496
Trp Ala Leu Ala Phe Phe Ser Phe Gly Val Phe Leu Leu Pro Gln Leu
470                 475                 480                 485

ttc ctc ttc acc atc tta aac tcg ctc tac ggt ttc ttc ctt ttc ctg      1544
Phe Leu Phe Thr Ile Leu Asn Ser Leu Tyr Gly Phe Phe Leu Phe Leu
                490                 495                 500

tgg ttc tgc tcc cag cgg tgc cgc tca gaa gca gag gcc aag gca cag      1592
Trp Phe Cys Ser Gln Arg Cys Arg Ser Glu Ala Glu Ala Lys Ala Gln
            505                 510                 515

ata gag gcc ttc agc tcc tcc caa aca aca cag tagtccgggc ctcctggcct    1645
Ile Glu Ala Phe Ser Ser Ser Gln Thr Thr Gln
        520                 525

ggaatcctca gcctctctgg ccgccagtag cctgaggcta cggctcctgc tagagagggt    1705
ggcaggcctg ctgctggacc ccagaggcca ctgtgaccgc caaggggcct tttccacttc    1765
cacggcctct ccaggcactg aggggaaggc attgctctac ctctccctga cattttgctc    1825
cggggcagat ccaaccttac ctggggcagc aaactttgtc ctggtacctg ggcccagctc    1885
gccagggatg tgggcagagc accagcctgg gcatcaggaa gccaagtttc aaggactgtc    1945
tttgagtctg tctgtatgac cttgggcctg ccacttctca cagaccctag gtatccacag    2005
ctgtgacatg gggcaagcg gctttgtttc agcctaaccc aggagcttag taaaaattgc     2065
ataagaccag ggggaagagt gtcagcgtgg ggtgggaatt cccgcggcct ccacctgctt    2125
gctaggggca ggatctcatt caggctgccc tggaagcacc tgcttggccc tgccaccttc    2185
ctccagggga gggccagatg gcatcctggc ttggggcggg tgggacctac ccaggctctg    2245
agactttact ggcctatgcc tgaggcctct tttcctttaa ctccctaaat tatgatgact    2305
ccaagtccaa gcccacccct cccaaagatt gggaggttcc gccgttccca gaggctcctc    2365
ctgcggtgct cccaagactt ccatagacca tctggaccag tagcccatcc cgcagttttc    2425
ttgggggcag aggaaaacgc ttctttctcc tccagctgaa tcagctggat cccagtgtcc    2485
tggctgtttg gtgattgggc aagattgaat ttgcccaggt aggcgtgaga gtgtgggttt    2545
taaattcgaa gctcaggcca tagtttcaga gaatcaccct taccccagac cttcatgaga    2605
cagtgctcat gaagccagtg cgtttcccag aacgaacact aggcggcacc gttggtccac    2665
actcagaggc ccttggcgcc aagactgcat ctagaatcgc tcaaacacct gtttgcagac    2725
cccatgcacc agctggaggg gccgtaactg caggactgcg cctactgagt gacccatttc    2785
ctccaggagg aaaggcaaga cacgcttaca cggccatttg tctcttttcc caatgcggcg    2845
gtgcactttc gctcttgggg gctgcacccc agacatagct ggcaccagag cagggtgctc    2905
aggtggtggg tgctcagggc cctgccccag gccactgggc cgttttgatg acctcgaagg    2965
tcacaggcag aaaataggag caggatttcc cctggggaaa agttctcctg ggacatcttc    3025
tgctcttctg tacatttcta gatgcaaata actccttcac caggcagtga gtggcgtagg    3085
ctctggagcc aggctgcctg ggctccaatg ccagctctgc cacttgctag ctgtgagact    3145
gtggacaaac cactcagcct ctgtgtgcct cagttttcct atttgtaaaa tagaggccat    3205
agtggtacct attttgaaga ctaagtaaaa gaattcaaat aaagagactt ggc           3258
```

<210> 29
<211> 528

<212> PRT
<213> Homo sapiens

<400> 29
Met Asp His Cys Gly Ala Leu Phe Leu Cys Leu Cys Leu Leu Thr Leu
1 5 10 15
Gln Asn Ala Thr Thr Glu Thr Trp Glu Glu Leu Leu Ser Tyr Met Glu
20 25 30
Asn Met Gln Val Ser Arg Gly Arg Ser Ser Val Phe Ser Ser Arg Gln
35 40 45
Leu His Gln Leu Glu Gln Met Leu Leu Asn Thr Ser Phe Pro Gly Tyr
50 55 60
Asn Leu Thr Leu Gln Thr Pro Thr Ile Gln Ser Leu Ala Phe Lys Leu
65 70 75 80
Ser Cys Asp Phe Ser Gly Leu Ser Leu Thr Ser Ala Thr Leu Lys Arg
85 90 95
Val Pro Gln Ala Gly Gly Gln His Ala Arg Gly Gln His Ala Met Gln
100 105 110
Phe Pro Ala Glu Leu Thr Arg Asp Ala Cys Lys Thr Arg Pro Arg Glu
115 120 125
Leu Arg Leu Ile Cys Ile Tyr Phe Ser Asn Thr His Phe Phe Lys Asp
130 135 140
Glu Asn Asn Ser Ser Leu Leu Asn Asn Tyr Val Leu Gly Ala Gln Leu
145 150 155 160
Ser His Gly His Val Asn Asn Leu Arg Asp Pro Val Asn Ile Ser Phe
165 170 175
Trp His Asn Gln Ser Leu Glu Gly Tyr Thr Leu Thr Cys Val Phe Trp
180 185 190
Lys Glu Gly Ala Arg Lys Gln Pro Trp Gly Gly Trp Ser Pro Glu Gly
195 200 205
Cys Arg Thr Glu Gln Pro Ser His Ser Gln Val Leu Cys Arg Cys Asn
210 215 220
His Leu Thr Tyr Phe Ala Val Leu Met Gln Leu Ser Pro Ala Leu Val
225 230 235 240
Pro Ala Glu Leu Leu Ala Pro Leu Thr Tyr Ile Ser Leu Val Gly Cys
245 250 255
Ser Ile Ser Ile Val Ala Ser Leu Ile Thr Val Leu Leu His Phe His
260 265 270
Phe Arg Lys Gln Ser Asp Ser Leu Thr Arg Ile His Met Asn Leu His
275 280 285
Ala Ser Val Leu Leu Leu Asn Ile Ala Phe Leu Leu Ser Pro Ala Phe
290 295 300
Ala Met Ser Pro Val Pro Gly Ser Ala Cys Thr Ala Leu Ala Ala Ala
305 310 315 320
Leu His Tyr Ala Leu Leu Ser Cys Leu Thr Trp Met Ala Ile Glu Gly
325 330 335
Phe Asn Leu Tyr Leu Leu Leu Gly Arg Val Tyr Asn Ile Tyr Ile Arg
340 345 350
Arg Tyr Val Phe Lys Leu Gly Val Leu Gly Trp Gly Ala Pro Ala Leu
355 360 365
Leu Val Leu Leu Ser Leu Ser Val Lys Ser Ser Val Tyr Gly Pro Cys
370 375 380
Thr Ile Pro Val Phe Asp Ser Trp Glu Asn Gly Thr Gly Phe Gln Asn
385 390 395 400
Met Ser Ile Cys Trp Val Arg Ser Pro Val Val His Ser Val Leu Val
405 410 415
Met Gly Tyr Gly Gly Leu Thr Ser Leu Phe Asn Leu Val Val Leu Ala
420 425 430

Trp Ala Leu Trp Thr Leu Arg Arg Leu Arg Glu Arg Ala Asp Ala Pro
435 440 445
Ser Val Arg Ala Cys His Asp Thr Val Thr Val Leu Gly Leu Thr Val
450 455 460
Leu Leu Gly Thr Thr Trp Ala Leu Ala Phe Phe Ser Phe Gly Val Phe
465 470 475 480
Leu Leu Pro Gln Leu Phe Leu Phe Thr Ile Leu Asn Ser Leu Tyr Gly
485 490 495
Phe Phe Leu Phe Leu Trp Phe Cys Ser Gln Arg Cys Arg Ser Glu Ala
500 505 510
Glu Ala Lys Ala Gln Ile Glu Ala Phe Ser Ser Ser Gln Thr Thr Gln
515 520 525

<210> 30
<211> 1584
<212> DNA
<213> Homo sapiens

<400> 30
atggatcact gtggtgccct tttcctgtgc ctgtgccttc tgactttgca gaatgcaaca 60
acagagacat gggaagaact cctgagctac atggagaata tgcaggtgtc caggggccgg 120
agctcagttt tttcctctcg tcaactccac cagctggagc agatgctact gaacaccagc 180
ttcccaggct acaacctgac cttgcagaca cccaccatcc agtctctggc cttcaagctg 240
agctgtgact tctctggcct ctcgctgacc agtgccactc tgaagcgggt gccccaggca 300
ggaggtcagc atgcccgggg tcagcacgcc atgcagttcc ccgccgagct gacccgggac 360
gcctgcaaga cccgccccag ggagctgcgg ctcatctgta tctacttctc caacacccac 420
tttttcaagg atgaaaacaa ctcatctctg ctgaataact acgtcctggg ggcccagctg 480
agtcatgggc acgtgaacaa cctcagggat cctgtgaaca tcagcttctg gcacaaccaa 540
agcctggaag gctacaccct gacctgtgtc ttctggaagg agggagccag gaaacagccc 600
tggggggggct ggagccctga gggctgtcgt acagagcagc cctcccactc tcaggtgctc 660
tgccgctgca accacctcac ctactttgct gttctcatgc aactctcccc agccctggtc 720
cctgcagagt tgctggcacc tcttacgtac atctccctcg tgggctgcag catctccatc 780
gtggcctcgc tgatcacagt cctgctgcac ttccatttca ggaagcagag tgactcctta 840
acacgcatcc acatgaacct gcatgcctcc gtgctgctcc tgaacatcgc cttcctgctg 900
agccccgcat tcgcaatgtc tcctgtgccc gggtcagcat gcacggctct ggccgctgcc 960
ctgcactacg cgctgctcag ctgcctcacc tggatggcca tcgagggctt caacctctac 1020
ctcctcctcg ggcgtgtcta caacatctac atccgcagat atgtgttcaa gcttggtgtg 1080
ctaggctggg gggccccagc cctcctggtg ctgctttccc tctctgtcaa gagctcggta 1140
tacggaccct gcacaatccc cgtcttcgac agctgggaga atggcacagg cttccagaac 1200
atgtccatat gctgggtgcg gagccccgtg gtgcacagtg tcctggtcat gggctacggc 1260
ggcctcacgt ccctcttcaa cctggtggtg ctggcctggg cgctgtggac cctgcgcagg 1320
ctgcgggagc gggcggatgc accaagtgtc agggcctgcc atgacactgt cactgtgctg 1380
ggcctcaccg tgctgctggg aaccacctgg gccttggcct tcttttcttt tggcgtcttc 1440
ctgctgcccc agctgttcct cttcaccatc ttaaactcgc tctacggttt cttccttttc 1500
ctgtggttct gctcccagcg gtgccgctca gaagcagagg ccaaggcaca gatagaggcc 1560
ttcagctcct cccaaacaac acag 1584

<210> 31
<211> 63
<212> PRT
<213> Homo sapiens

<400> 31
Leu Lys Ser Pro Glu Gly Lys Ser Arg Lys Asn Pro Ala Arg Thr Cys
1 5 10 15
Lys Asp Leu Phe Leu Cys His Pro Glu Phe Lys Ser Gly Glu Tyr Trp
20 25 30
Ile Asp Pro Asn Gln Gly Cys Ile Lys Asp Ala Ile Lys Val Phe Cys

```
        35                  40                  45
Asn Lys Arg Phe Glu Thr Gly Val Gly Glu Thr Cys Ile Ser Pro
    50                  55                  60

      <210> 32
      <211> 25
      <212> PRT
      <213> Homo sapiens

      <400> 32
Ile Ser Asn Val Gln Thr Phe Leu Arg Leu Leu Ser Thr Glu Ala Ser
 1               5                  10                  15
Gln Asn Ile Thr Tyr His Cys Lys Asn
            20                  25

      <210> 33
      <211> 33
      <212> PRT
      <213> Homo sapiens

      <400> 33
Thr Val Leu Gly Glu Asp Gly Cys Ser Ser Arg Thr Gly Glu Trp Gly
 1               5                  10                  15
Lys Thr Val Ile Glu Tyr Glu Thr Lys Lys Thr Thr Arg Leu Pro Ile
            20                  25                  30
Val


      <210> 34
      <211> 65
      <212> PRT
      <213> Homo sapiens

      <400> 34
Ile Asn Thr Ile Lys Asn Pro Leu Gly Thr Arg Asp Asn Pro Ala Arg
 1               5                  10                  15
Ile Cys Lys Asp Leu Leu Asn Cys Glu Gln Lys Val Ser Asp Gly Lys
            20                  25                  30
Tyr Trp Ile Asp Pro Asn Leu Gly Cys Pro Ser Asp Ala Ile Glu Val
        35                  40                  45
Phe Ile Asn Thr Cys Asn Phe Ser Ala Gly Gly Gln Thr Cys Leu Pro
    50                  55                  60
Pro
65

      <210> 35
      <211> 26
      <212> PRT
      <213> Homo sapiens

      <400> 35
Val Gly Lys Val Gln Met Asn Phe Leu His Leu Leu Ser Ser Glu Ala
 1               5                  10                  15
Thr His Ile Ile Thr Ile His Cys Leu Asn
            20                  25

      <210> 36
      <211> 32
```

<212> PRT
<213> Homo sapiens

<400> 36
Lys Val Leu Ser Asp Asp Cys Lys Ile Gln Asp Gly Ser Trp His Lys
1 5 10 15
Ala Thr Phe Leu Phe His Thr Gln Glu Pro Asn Gln Leu Pro Val Ile
20 25 30

<210> 37
<211> 31
<212> PRT
<213> Homo sapiens

<400> 37
Gly Glu Ser Val Thr Leu Thr Cys Ser Val Ser Gly Phe Gly Pro Pro
1 5 10 15
Pro Val Thr Trp Leu Arg Asn Gly Lys Leu Ser Leu Thr Ile Ser
20 25 30

<210> 38
<211> 57
<212> PRT
<213> Homo sapiens

<400> 38
Gly Arg Thr Val Arg Leu Gln Cys Pro Val Glu Gly Asp Pro Pro Pro
1 5 10 15
Thr Met Trp Thr Lys Asp Gly Arg Thr Ile His Ser Gly Trp Ser Arg
20 25 30
Phe Arg Val Leu Pro Gln Gly Leu Lys Val Lys Gln Val Glu Arg Glu
35 40 45
Asp Ala Gly Val Tyr Val Cys Lys Ala
50 55

<210> 39
<211> 59
<212> PRT
<213> Homo sapiens

<400> 39
Gly Ser Ser Val Arg Leu Lys Cys Val Ala Ser Gly His Pro Arg Pro
1 5 10 15
Asp Ile Thr Trp Met Lys Asp Asp Gln Ala Leu Thr Arg Pro Glu Ala
20 25 30
Ala Glu Pro Arg Lys Lys Lys Trp Thr Leu Ser Leu Lys Asn Leu Arg
35 40 45
Pro Glu Asp Ser Gly Lys Tyr Thr Cys Arg Val
50 55

<210> 40
<211> 79
<212> PRT
<213> Homo sapiens

<400> 40
Gly Gly Thr Thr Ser Phe Gln Cys Lys Val Arg Ser Asp Val Lys Pro
1 5 10 15

Val Ile Gln Trp Leu Lys Arg Val Glu Tyr Gly Ala Glu Gly Arg His
20 25 30
Asn Ser Thr Ile Asp Val Gly Gly Gln Lys Phe Val Val Leu Pro Thr
35 40 45
Gly Asp Val Trp Ser Arg Pro Asp Gly Ser Tyr Asn Lys Leu Leu Ile
50 55 60
Thr Arg Ala Arg Gln Asp Asp Ala Gly Met Tyr Ile Cys Leu Gly
65 70 75

<210> 41
<211> 78
<212> PRT
<213> Homo sapiens

<400> 41
Arg Gly Ser Leu Thr Val Gln Cys Val Tyr Arg Ser Gly Trp Glu Thr
1 5 10 15
Tyr Leu Lys Trp Trp Cys Arg Gly Ala Ile Trp Arg Asp Cys Lys Ile
20 25 30
Leu Val Lys Thr Ser Gly Ser Glu Gln Glu Val Lys Arg Asp Arg Val
35 40 45
Ser Ile Lys Asp Asn Gln Lys Asn Arg Thr Phe Thr Val Thr Met Glu
50 55 60
Asp Leu Met Lys Thr Asp Ala Asp Thr Tyr Trp Cys Gly Ile
65 70 75

<210> 42
<211> 10
<212> PRT
<213> Homo sapiens

<400> 42
Val Phe Val Leu Gly Thr Leu Gly Ile Phe
1 5 10

<210> 43
<211> 10
<212> PRT
<213> Homo sapiens

<400> 43
Val Phe Ile Leu Gly Thr Leu Leu Leu Trp
1 5 10

<210> 44
<211> 116
<212> PRT
<213> Homo sapiens

<400> 44
Cys Gly Gly Thr Leu Asp Leu Thr Glu Ser Ser Gly Ser Ile Ser Ser
1 5 10 15
Pro Asn Tyr Pro Asn Arg Ser Asp Tyr Pro Pro Asn Lys Glu Cys Val
20 25 30
Trp Arg Ile Arg Ala Pro Pro Gly Tyr Arg Val Val Glu Leu Thr Phe
35 40 45
Gln Asp Phe Asp Leu Glu Asp His Asp Gly Ala Pro Cys Arg Tyr Asp
50 55 60

```
Tyr Val Glu Ile Arg Asp Gly Asp Pro Ser Ser Pro Leu Leu Gly Arg
65                  70                  75                  80
Phe Cys Gly Ser Gly Lys Pro Glu Asp Ile Arg Ser Thr Ser Asn Arg
                85                  90                  95
Met Leu Ile Lys Phe Val Ser Asp Ala Ser Val Ser Lys Arg Gly Phe
            100                 105                 110
Lys Ala Thr Tyr
        115

      <210> 45
      <211> 97
      <212> PRT
      <213> Homo sapiens

      <400> 45
Gly Ser Val Leu Leu Ala Gln Glu Leu Pro Gln Gln Leu Thr Ser Pro
 1               5                  10                  15
Gly Tyr Pro Glu Pro Tyr Gly Lys Gly Gln Glu Ser Ser Thr Asp Ile
            20                  25                  30
Lys Ala Pro Glu Gly Phe Ala Val Arg Leu Val Phe Gln Asp Phe Asp
        35                  40                  45
Leu Glu Pro Ser Gln Asp Cys Ala Gly Asp Ser Val Thr Val Ser Trp
    50                  55                  60
Gly Trp Gly Gly Ser Arg Gln Asp Cys Gly Gln Gly Asp Ser Arg Gly
65                  70                  75                  80
Cys Gly Lys Trp Arg Cys Pro Glu Ser Pro Ile Trp Arg Arg Asp Glu
                85                  90                  95
Phe


      <210> 46
      <211> 45
      <212> PRT
      <213> Homo sapiens

      <400> 46
Cys Ala Pro Asn Asn Pro Cys Ser Asn Gly Gly Thr Cys Val Asn Thr
 1               5                  10                  15
Pro Gly Gly Ser Ser Asp Asn Phe Gly Gly Tyr Thr Cys Glu Cys Pro
            20                  25                  30
Pro Gly Asp Tyr Tyr Leu Ser Tyr Thr Gly Lys Arg Cys
        35                  40                  45

      <210> 47
      <211> 67
      <212> PRT
      <213> Homo sapiens

      <400> 47
Trp Ser Thr Asp Lys His Ile Gly Gly Arg Thr Ser Leu Gly Phe Asn
 1               5                  10                  15
Leu Glu Tyr Arg Ile Arg Val Thr Cys Asp Glu Asn Tyr Tyr Gly Glu
            20                  25                  30
Gly Cys Asn Lys Phe Cys Arg Pro Arg Asp Asp Ala Phe Gly His Tyr
        35                  40                  45
Thr Cys Asp Glu Asn Gly Asn Lys Leu Cys Leu Glu Gly Trp Lys Gly
    50                  55                  60
Glu Tyr Cys
```

65

<210> 48
<211> 59
<212> PRT
<213> Homo sapiens

<400> 48
Cys Asp Cys Asn Pro His Gly Ser Leu Ser Asp Asp Thr Cys Asp Ser
1 5 10 15
Asp Asp Glu Leu Phe Gly Glu Glu Thr Gly Gln Cys Leu Lys Cys Lys
20 25 30
Pro Asn Val Thr Gly Arg Arg Cys Asp Arg Cys Lys Pro Gly Tyr Tyr
35 40 45
Gly Leu Pro Ser Gly Asp Pro Gln Gln Gly Cys
50 55

<210> 49
<211> 31
<212> PRT
<213> Homo sapiens

<400> 49
Cys Val Pro Leu Cys Ala Gln Glu Cys Val His Gly Arg Cys Val Ala
1 5 10 15
Pro Asn Gln Cys Gln Cys Val Pro Gly Trp Arg Gly Asp Asp Cys
20 25 30

<210> 50
<211> 30
<212> PRT
<213> Homo sapiens

<400> 50
Cys Gln Phe Arg Cys Gln Cys His Gly Ala Pro Cys Asp Pro Gln Thr
1 5 10 15
Gly Ala Cys Phe Cys Pro Ala Glu Arg Thr Gly Pro Ser Cys
20 25 30

<210> 51
<211> 31
<212> PRT
<213> Homo sapiens

<400> 51
Cys Pro Ser Thr His Pro Cys Gln Asn Gly Gly Val Phe Gln Thr Pro
1 5 10 15
Gln Gly Ser Cys Ser Cys Pro Pro Gly Trp Met Gly Thr Ile Cys
20 25 30

<210> 52
<211> 31
<212> PRT
<213> Homo sapiens

<400> 52
Cys Ser Gln Glu Cys Arg Cys His Asn Gly Gly Leu Cys Asp Arg Phe
1 5 10 15

Thr Gly Gln Cys Arg Cys Ala Pro Gly Tyr Thr Gly Asp Arg Cys
20 25 30

<210> 53
<211> 31
<212> PRT
<213> Homo sapiens

<400> 53
Cys Ala Glu Thr Cys Asp Cys Ala Pro Asp Ala Arg Cys Phe Pro Ala
1 5 10 15
Asn Gly Ala Cys Leu Cys Glu His Gly Phe Thr Gly Asp Arg Cys
20 25 30

<210> 54
<211> 27
<212> PRT
<213> Homo sapiens

<400> 54
Cys Asp Arg Glu His Ser Leu Ser Cys His Pro Met Asn Gly Glu Cys
1 5 10 15
Ser Cys Leu Pro Gly Trp Ala Gly Leu His Cys
20 25

<210> 55
<211> 31
<212> PRT
<213> Homo sapiens

<400> 55
Cys Gln Glu His Cys Leu Cys Leu His Gly Gly Val Cys Gln Ala Thr
1 5 10 15
Ser Gly Leu Cys Gln Cys Ala Pro Gly Tyr Thr Gly Pro His Cys
20 25 30

<210> 56
<211> 31
<212> PRT
<213> Homo sapiens

<400> 56
Cys Ser Ala Arg Cys Ser Cys Glu Asn Ala Ile Ala Cys Ser Pro Ile
1 5 10 15
Asp Gly Glu Cys Val Cys Lys Glu Gly Trp Gln Arg Gly Asn Cys
20 25 30

<210> 57
<211> 31
<212> PRT
<213> Homo sapiens

<400> 57
Cys Asn Ala Ser Cys Gln Cys Ala His Glu Ala Val Cys Ser Pro Gln
1 5 10 15
Thr Gly Ala Cys Thr Cys Thr Pro Gly Trp His Gly Ala His Cys
20 25 30

<210> 58
<211> 31
<212> PRT
<213> Homo sapiens

<400> 58
Cys Ala Ser Arg Cys Asp Cys Asp His Ser Asp Gly Cys Asp Pro Val
1 5 10 15
His Gly Arg Cys Gln Cys Gln Ala Gly Trp Met Gly Ala Arg Cys
20 25 30

<210> 59
<211> 31
<212> PRT
<213> Homo sapiens

<400> 59
Cys Ser Asn Thr Cys Thr Cys Lys Asn Gly Gly Thr Cys Leu Pro Glu
1 5 10 15
Asn Gly Asn Cys Val Cys Ala Pro Gly Phe Arg Gly Pro Ser Cys
20 25 30

<210> 60
<211> 30
<212> PRT
<213> Homo sapiens

<400> 60
Cys Val Pro Cys Lys Cys Ala Asn His Ser Phe Cys His Pro Ser Asn
1 5 10 15
Gly Thr Cys Tyr Cys Leu Ala Gly Trp Thr Gly Pro Asp Cys
20 25 30

<210> 61
<211> 31
<212> PRT
<213> Homo sapiens

<400> 61
Cys Ala Gln Thr Cys Gln Cys His His Gly Gly Thr Cys His Pro Gln
1 5 10 15
Asp Gly Ser Cys Ile Cys Pro Leu Gly Trp Thr Gly His His Cys
20 25 30

<210> 62
<211> 31
<212> PRT
<213> Homo sapiens

<400> 62
Cys Ser Gln Pro Cys Gln Cys Gly Pro Gly Glu Lys Cys His Pro Glu
1 5 10 15
Thr Gly Ala Cys Val Cys Pro Pro Gly His Ser Gly Ala Pro Cys
20 25 30

<210> 63
<211> 37
<212> PRT

<213> Homo sapiens

<400> 63
Gln Thr Gly Ala Cys Thr Cys Thr Pro Gly Trp His Gly Ala His Cys
1 5 10 15
Gln Leu Pro Cys Pro Lys Gly Gln Phe Gly Glu Gly Cys Ala Ser Arg
20 25 30
Cys Asp Cys Asp His
35

<210> 64
<211> 31
<212> PRT
<213> Mus musculus

<400> 64
Cys Ser Asn Thr Cys Thr Cys Lys Asn Gly Gly Thr Cys Val Ser Glu
1 5 10 15
Asn Gly Asn Cys Val Cys Ala Pro Gly Phe Arg Gly Pro Ser Cys
20 25 30

<210> 65
<211> 31
<212> PRT
<213> Mus musculus

<400> 65
Cys Val Gln Cys Lys Cys Asn Asn Asn His Ser Ser Cys His Pro Ser
1 5 10 15
Asp Gly Thr Cys Ser Cys Leu Ala Gly Trp Thr Gly Pro Asp Cys
20 25 30

<210> 66
<211> 31
<212> PRT
<213> Mus musculus

<400> 66
Cys Ser Gln Leu Cys Gln Cys His His Gly Gly Thr Cys His Pro Gln
1 5 10 15
Asp Gly Ser Cys Ile Cys Thr Pro Gly Trp Thr Gly Pro Asn Cys
20 25 30

<210> 67
<211> 31
<212> PRT
<213> Mus musculus

<400> 67
Cys Ser Gln Leu Cys Gln Cys Asp Leu Gly Glu Met Cys His Pro Glu
1 5 10 15
Thr Gly Ala Cys Val Cys Pro Pro Gly His Ser Gly Ala Asp Cys
20 25 30

<210> 68
<211> 35
<212> PRT
<213> Mus musculus

<400> 68
His Ala Ser Gly Asp Pro Val His Gly Gln Cys Arg Cys Gln Ala Gly
1 5 10 15
Trp Met Gly Thr Arg Cys His Leu Pro Cys Pro Glu Gly Phe Trp Gly
20 25 30
Ala Asn Cys
35

<210> 69
<211> 40
<212> PRT
<213> Mus musculus

<400> 69
Cys Thr Cys Lys Asn Gly Gly Thr Cys Val Ser Glu Asn Gly Asn Cys
1 5 10 15
Val Cys Ala Pro Gly Phe Arg Gly Pro Ser Cys Gln Arg Pro Cys Pro
20 25 30
Pro Gly Arg Tyr Gly Lys Arg Cys
35 40

<210> 70
<211> 35
<212> PRT
<213> Mus musculus

<400> 70
Cys Lys Cys Asn Asn Asn His Ser Ser Cys His Pro Ser Asp Gly Thr
1 5 10 15
Cys Ser Cys Leu Ala Gly Trp Thr Gly Pro Asp Cys Ser Glu Ala Cys
20 25 30
Pro Pro Gly
35

<210> 71
<211> 34
<212> PRT
<213> Mus musculus

<400> 71
Cys Gln Cys His His Gly Gly Thr Cys His Pro Gln Asp Gly Ser Cys
1 5 10 15
Ile Cys Thr Pro Gly Trp Thr Gly Pro Asn Cys Leu Glu Gly Cys Pro
20 25 30
Pro Arg

<210> 72
<211> 58
<212> PRT
<213> Mus musculus

<400> 72
His Gly Gln Cys Arg Cys Gln Ala Gly Trp Met Gly Thr Arg Cys His
1 5 10 15
Leu Pro Cys Pro Glu Gly Phe Trp Gly Ala Asn Cys Ser Asn Thr Cys
20 25 30
Thr Cys Lys Asn Gly Gly Thr Cys Val Ser Glu Asn Gly Asn Cys Val

35 40 45
Cys Ala Pro Gly Phe Arg Gly Pro Ser Cys
50 55

<210> 73
<211> 28
<212> PRT
<213> Rattus sp.

<400> 73
Glu Cys Arg Cys His Asn Gly Gly Leu Cys Asp Arg Phe Thr Gly Gln
1 5 10 15
Cys His Cys Ala Pro Gly Tyr Ile Gly Asp Arg Cys
20 25

<210> 74
<211> 31
<212> PRT
<213> Rattus sp.

<400> 74
Cys Ala Glu Thr Cys Asp Cys Ala Pro Gly Ala Arg Cys Phe Pro Ala
1 5 10 15
Asn Gly Ala Cys Leu Cys Glu His Gly Phe Thr Gly Asp Arg Cys
20 25 30

<210> 75
<211> 33
<212> PRT
<213> Rattus sp.

<400> 75
Cys Gln Asp Pro Cys Thr Cys Asp Pro Glu His Ser Leu Ser Cys His
1 5 10 15
Pro Met His Gly Glu Cys Ser Cys Gln Pro Gly Trp Ala Gly Leu His
20 25 30
Cys

<210> 76
<211> 31
<212> PRT
<213> Rattus sp.

<400> 76
Cys Gln Glu His Cys Leu Cys Leu His Gly Gly Val Cys Leu Ala Asp
1 5 10 15
Ser Gly Leu Cys Arg Cys Ala Pro Gly Tyr Thr Gly Pro His Cys
20 25 30

<210> 77
<211> 31
<212> PRT
<213> Rattus sp.

<400> 77
Cys Ser Ser His Cys Ser Cys Glu Asn Ala Ile Ala Cys Ser Pro Val
1 5 10 15

Asp Gly Thr Cys Ile Cys Lys Glu Gly Trp Gln Arg Gly Asn Cys
20 25 30

<210> 78
<211> 31
<212> PRT
<213> Rattus sp.

<400> 78
Cys Asn Ala Ser Cys Gln Cys Ala His Glu Gly Val Cys Ser Pro Gln
1 5 10 15
Thr Gly Ala Cys Thr Cys Thr Pro Gly Trp Arg Gly Val His Cys
20 25 30

<210> 79
<211> 31
<212> PRT
<213> Rattus sp.

<400> 79
Cys Ala Ser Val Cys Asp Cys Asp His Ser Asp Gly Cys Asp Pro Val
1 5 10 15
His Gly His Cys Arg Cys Gln Ala Gly Trp Met Gly Thr Arg Cys
20 25 30

<210> 80
<211> 31
<212> PRT
<213> Rattus sp.

<400> 80
Cys Ser Asn Ala Cys Thr Cys Lys Asn Gly Gly Thr Cys Val Pro Glu
1 5 10 15
Asn Gly Asn Cys Val Cys Ala Pro Gly Phe Arg Gly Pro Ser Cys
20 25 30

<210> 81
<211> 30
<212> PRT
<213> Rattus sp.

<400> 81
Cys Val Pro Cys Lys Cys Asn Asn His Ser Ser Cys His Pro Ser Asp
1 5 10 15
Gly Thr Cys Ser Cys Leu Ala Gly Trp Thr Gly Pro Asp Cys
20 25 30

<210> 82
<211> 31
<212> PRT
<213> Rattus sp.

<400> 82
Cys Ser Gln Pro Cys Gln Cys His His Gly Ala Thr Cys His Pro Gln
1 5 10 15
Asp Gly Ser Cys Val Cys Ile Pro Gly Trp Thr Gly Pro Asn Cys
20 25 30

<210> 83
<211> 31
<212> PRT
<213> Rattus sp.

<400> 83
Cys Ser Gln Leu Cys Gln Cys Asp Pro Gly Glu Met Cys His Pro Glu
1 5 10 15
Thr Gly Ala Cys Val Cys Pro Pro Gly His Ser Gly Ala His Cys
20 25 30

<210> 84
<211> 40
<212> PRT
<213> Rattus sp.

<400> 84
Cys Arg Cys His Asn Gly Gly Leu Cys Asp Arg Phe Thr Gly Gln Cys
1 5 10 15
His Cys Ala Pro Gly Tyr Ile Gly Asp Arg Cys Arg Glu Glu Cys Pro
20 25 30
Val Gly Arg Phe Gly Gln Asp Cys
35 40

<210> 85
<211> 39
<212> PRT
<213> Rattus sp.

<400> 85
Cys Asp Cys Ala Pro Gly Ala Arg Cys Phe Pro Ala Asn Gly Ala Cys
1 5 10 15
Leu Cys Glu His Gly Phe Thr Gly Asp Arg Cys Thr Glu Arg Leu Cys
20 25 30
Pro Asp Gly Tyr Gly Leu Cys
35

<210> 86
<211> 42
<212> PRT
<213> Rattus sp.

<400> 86
Cys Thr Cys Asp Pro Glu His Ser Leu Ser Cys His Pro Met His Gly
1 5 10 15
Glu Cys Ser Cys Gln Pro Gly Trp Ala Gly Leu His Cys Asn Glu Ser
20 25 30
Cys Pro Gln Asp Thr His Gly Ala Gly Cys
35 40

<210> 87
<211> 40
<212> PRT
<213> Rattus sp.

<400> 87
Cys Leu Cys Leu His Gly Gly Val Cys Leu Ala Asp Ser Gly Leu Cys
1 5 10 15

```
Arg Cys Ala Pro Gly Tyr Thr Gly Pro His Cys Ala Asn Leu Cys Pro
            20                  25                  30
Pro Asn Thr Tyr Gly Ile Asn Cys
        35                  40

      <210> 88
      <211> 40
      <212> PRT
      <213> Rattus sp.

      <400> 88
Cys Ser Cys Glu Asn Ala Ile Ala Cys Ser Pro Val Asp Gly Thr Cys
 1               5                  10                  15
Ile Cys Lys Glu Gly Trp Gln Arg Gly Asn Cys Ser Val Pro Cys Pro
            20                  25                  30
Pro Gly Thr Trp Gly Phe Ser Cys
        35                  40

      <210> 89
      <211> 40
      <212> PRT
      <213> Rattus sp.

      <400> 89
Cys Gln Cys Ala His Glu Gly Val Cys Ser Pro Gln Thr Gly Ala Cys
 1               5                  10                  15
Thr Cys Thr Pro Gly Trp Arg Gly Val His Cys Gln Leu Pro Cys Pro
            20                  25                  30
Lys Gly Gln Phe Gly Glu Gly Cys
        35                  40

      <210> 90
      <211> 40
      <212> PRT
      <213> Rattus sp.

      <400> 90
Cys Asp Cys Asp His Ser Asp Gly Cys Asp Pro Val His Gly His Cys
 1               5                  10                  15
Arg Cys Gln Ala Gly Trp Met Gly Thr Arg Cys His Leu Pro Cys Pro
            20                  25                  30
Glu Gly Phe Trp Gly Ala Asn Cys
        35                  40

      <210> 91
      <211> 40
      <212> PRT
      <213> Rattus sp.

      <400> 91
Cys Thr Cys Lys Asn Gly Gly Thr Cys Val Pro Glu Asn Gly Asn Cys
 1               5                  10                  15
Val Cys Ala Pro Gly Phe Arg Gly Pro Ser Cys Gln Arg Pro Cys Pro
            20                  25                  30
Pro Gly Arg Tyr Gly Lys Arg Cys
        35                  40

      <210> 92
```

<211> 40
<212> PRT
<213> Rattus sp.

<400> 92
Cys Lys Cys Asn Asn His Ser Ser Cys His Pro Ser Asp Gly Thr Cys
1 5 10 15
Ser Cys Leu Ala Gly Trp Thr Gly Pro Asp Cys Ser Glu Ser Cys Pro
20 25 30
Pro Gly His Trp Gly Leu Lys Cys
35 40

<210> 93
<211> 40
<212> PRT
<213> Rattus sp.

<400> 93
Cys Gln Cys His His Gly Ala Thr Cys His Pro Gln Asp Gly Ser Cys
1 5 10 15
Val Cys Ile Pro Gly Trp Thr Gly Pro Asn Cys Ser Glu Gly Cys Pro
20 25 30
Ser Arg Met Phe Gly Val Asn Cys
35 40

<210> 94
<211> 36
<212> PRT
<213> Rattus sp.

<400> 94
Cys Gln Cys Asp Pro Gly Glu Met Cys His Pro Glu Thr Gly Ala Cys
1 5 10 15
Val Cys Pro Pro Gly His Ser Gly Ala His Cys Lys Val Gly Ser Gln
20 25 30
Glu Ser Phe Thr
35

<210> 95
<211> 64
<212> PRT
<213> Rattus sp.

<400> 95
Gly Val Cys Ser Pro Gln Thr Gly Ala Cys Thr Cys Thr Pro Gly Trp
1 5 10 15
Arg Gly Val His Cys Gln Leu Pro Cys Pro Lys Gly Gln Phe Gly Glu
20 25 30
Gly Cys Ala Ser Val Cys Asp Cys Asp His Ser Asp Gly Cys Asp Pro
35 40 45
Val His Gly His Cys Arg Cys Gln Ala Gly Trp Met Gly Thr Arg Cys
50 55 60

<210> 96
<211> 129
<212> PRT
<213> Homo sapiens

<400> 96

```
Gln Glu Ser Arg Ala Gln Lys Phe Leu Arg Gln His Ile Asp Ser Pro
 1               5                  10                  15
Lys Thr Ser Ser Ser Asn Pro Asn Tyr Cys Asn Gln Met Met Asp Lys
            20                  25                  30
Arg Arg Asn Met Thr Gln Gln Arg Cys Lys Pro Val Asn Thr Phe Val
        35                  40                  45
His Glu Ser Leu Ala Asp Val Lys Ala Val Cys Ser Gln Lys Asn Val
    50                  55                  60
Thr Cys Lys Asn Gly Gln Ser Lys Ser Ser Phe Gln Ile Thr Asp Cys
65                  70                  75                  80
Arg Leu Thr Gly Gly Ser Gln Lys Tyr Pro Asn Cys Arg Tyr Arg Thr
                85                  90                  95
Ser Ala Ser Thr Lys His Ile Ile Val Ala Cys Glu Gly Arg Asp Arg
            100                 105                 110
Asp Asp Pro Tyr Tyr Asn Pro Tyr Val Pro Val His Phe Asp Ala Ser
        115                 120                 125
Val
```

<210> 97
<211> 125
<212> PRT
<213> Homo sapiens

<400> 97

```
Gly Met Thr Ser Ser Gln Trp Phe Lys Ile Gln His Met Gln Pro Ser
 1               5                  10                  15
Pro Gln Ala Cys Asn Ser Ala Met Lys Asn Ile Asn Lys His Thr Lys
            20                  25                  30
Arg Cys Lys Asp Leu Asn Thr Phe Leu His Glu Pro Phe Ser Ser Val
        35                  40                  45
Ala Ala Thr Cys Gln Thr Pro Lys Ile Ala Cys Lys Asn Gly Asp Lys
    50                  55                  60
Asn Cys His Gln Ser His Gly Pro Val Ser Leu Thr Met Cys Lys Leu
65                  70                  75                  80
Thr Ser Gly Lys Tyr Pro Asn Cys Arg Tyr Lys Glu Lys Arg Gln Asn
                85                  90                  95
Lys Ser Tyr Val Val Ala Cys Lys Pro Pro Gln Lys Lys Asp Ser Gln
            100                 105                 110
Gln Phe His Leu Val Pro Val His Leu Asp Arg Val Leu
        115                 120                 125
```

<210> 98
<211> 411
<212> PRT
<213> Homo sapiens

<400> 98

```
Cys Asn Arg Thr Trp Asp Gly Ile Thr Cys Trp Pro Asp Thr Pro Pro
 1               5                  10                  15
Gly Glu Leu Val Val Val Pro Cys Pro Lys Tyr Phe Tyr Gly Phe Ser
            20                  25                  30
Ser Asp Gln Thr Asp Thr Thr Gly Asn Val Ser Arg Asn Cys Thr Glu
        35                  40                  45
Asp Gly Ser Trp Ser Glu Pro Pro Pro Ser Asn Arg Thr Trp Arg Asn
    50                  55                  60
Tyr Ser Ala Cys Gly Glu Asp Asp Pro Glu Glu Glu Ser Glu Lys Lys
```

```
65                  70                      75                  80
Lys Lys Tyr Tyr Leu Val Leu Lys Ile Ile Tyr Thr Val Gly Tyr Ser
                85                      90                  95
Leu Ser Leu Ala Ala Leu Leu Val Ala Val Val Ile Leu Leu Leu Phe
            100                     105                 110
Arg Lys Leu His Thr Leu Trp Pro Asp Asn Ala Asp Gly Ala Leu Glu
        115                 120                     125
Val Gly Ala Pro Trp Gly Ala Pro Phe Gln Val Arg Arg Ser Ile Arg
    130                 135                     140
Cys Thr Arg Asn Tyr Ile His Met Asn Leu Phe Leu Ser Phe Ile Leu
145                 150                     155                 160
Arg Ala Ala Ser Val Phe Ile Lys Asp Ala Val Leu Lys Ser Glu Val
                165                     170                 175
Ser Ser Asp Glu Pro Glu Arg Leu Ser Ser Arg Cys Ser Leu Ser Thr
            180                     185                 190
Gly Gln Val Val Val Gly Cys Lys Leu Leu Val Val Phe Gln Phe Gln
        195                 200                     205
Tyr Cys Val Met Thr Asn Phe Phe Trp Leu Leu Val Glu Gly Leu Tyr
    210                 215                     220
Leu His Thr Leu Leu Val Val Thr Phe Phe Ser Glu Arg Lys Tyr Leu
225                 230                     235                 240
Trp Trp Tyr Leu Leu Ile Gly Trp Gly Val Pro Leu Val Phe Val Thr
                245                     250                 255
Val Trp Ala Ile Val Arg Leu Leu Phe Glu Asp Thr Gly Cys Trp Asp
            260                     265                 270
Ser Asn Gly Leu Ala Met Phe Pro Glu Ala Lys Met Cys Ile Trp Met
        275                 280                     285
Ser Asp Asn Ser His Leu Trp Trp Ile Ile Lys Gly Pro Ile Leu Leu
    290                 295                     300
Ser Ile Leu Val Asn Phe Phe Leu Phe Ile Asn Ile Ile Arg Ile Leu
305                 310                     315                 320
Val Thr Lys Leu Arg Ala Ala Gln Thr Gly Glu Thr Asp Gln Arg Gln
                325                     330                 335
Tyr Ser Gln Tyr Arg Lys Leu Ala Lys Ser Thr Leu Leu Leu Ile Pro
            340                     345                 350
Leu Phe Gly Ile His Tyr Val Val Phe Ala Phe Arg Pro Ser Asn Asp
        355                 360                     365
Ala Arg Gly Val Leu Arg Lys Ile Lys Leu Tyr Phe Glu Leu Ser Leu
    370                 375                     380
Gly Ser Phe Gln Gly Phe Phe Val Ala Val Leu Tyr Cys Phe Leu Asn
385                 390                     395                 400
Gly Glu Val Gln Ala Glu Ile Arg Arg Arg Trp
                405                     410
```

```
<210> 99
<211> 328
<212> PRT
<213> Homo sapiens

<400> 99
Leu Thr Cys Val Phe Trp Lys Glu Gly Ala Arg Lys Gln Pro Trp Gly
 1               5                  10                  15
Gly Trp Ser Pro Glu Gly Cys Arg Thr Glu Gln Pro Ser His Ser Gln
            20                  25                  30
Val Leu Cys Arg Cys Asn His Leu Thr Tyr Phe Ala Val Leu Met Gln
        35                  40                  45
Leu Ser Pro Ala Leu Val Pro Ala Glu Leu Leu Ala Pro Leu Thr Tyr
    50                  55                  60
```

Ile Ser Leu Val Gly Cys Ser Ile Ser Ile Val Ala Ser Leu Ile Thr
65 70 75 80
Val Leu Leu His Phe Arg Lys Gln Ser Asp Ser Leu Thr Arg Ile His
85 90 95
Met Asn Leu His Ala Ser Val Leu Leu Leu Asn Ile Ala Phe Leu Leu
100 105 110
Ser Pro Ala Phe Ala Met Ser Pro Val Pro Gly Ser Ala Cys Thr Ala
115 120 125
Leu Ala Ala Ala Leu His Tyr Ala Leu Leu Ser Cys Leu Thr Trp Met
130 135 140
Ala Ile Glu Gly Phe Asn Leu Tyr Leu Leu Leu Gly Arg Val Tyr Asn
145 150 155 160
Ile Tyr Ile Arg Arg Tyr Val Phe Lys Leu Gly Val Leu Gly Trp Gly
165 170 175
Ala Pro Ala Leu Leu Val Leu Leu Ser Leu Ser Val Lys Ser Ser Val
180 185 190
Tyr Gly Pro Cys Thr Ile Pro Val Phe Asp Ser Trp Glu Asn Gly Thr
195 200 205
Gly Phe Gln Asn Met Ser Ile Cys Trp Val Arg Ser Pro Val Val His
210 215 220
Ser Val Leu Val Met Gly Tyr Gly Gly Leu Thr Ser Leu Phe Asn Leu
225 230 235 240
Val Val Leu Ala Trp Ala Leu Trp Thr Leu Arg Arg Leu Arg Glu Arg
245 250 255
Ala Asp Ala Pro Ser Val Arg Ala Cys His Asp Thr Val Thr Val Leu
260 265 270
Gly Leu Thr Val Leu Leu Gly Thr Thr Trp Ala Leu Ala Phe Phe Ser
275 280 285
Phe Gly Val Phe Leu Leu Pro Gln Leu Phe Leu Phe Thr Ile Leu Asn
290 295 300
Ser Leu Tyr Gly Phe Phe Leu Phe Leu Trp Phe Cys Ser Gln Arg Cys
305 310 315 320
Arg Ser Glu Ala Glu Ala Lys Ala
325

<210> 100
<211> 150
<212> PRT
<213> Pan troglodytes

<400> 100
Met Val Leu Cys Phe Pro Leu Leu Leu Leu Leu Leu Val Leu Trp Gly
1 5 10 15
Pro Val Cys Pro Leu His Ala Trp Pro Lys Arg Leu Thr Lys Ala His
20 25 30
Trp Phe Glu Ile Gln His Ile Gln Pro Ser Pro Leu Gln Cys Asn Arg
35 40 45
Ala Met Ser Gly Ile Asn Asn Tyr Ala Gln His Cys Lys His Gln Asn
50 55 60
Thr Phe Leu His Asp Ser Phe Gln Asn Val Ala Ala Val Cys Asp Leu
65 70 75 80
Leu Ser Ile Val Cys Lys Asn Arg Arg His Asn Cys His Gln Ser Ser
85 90 95
Lys Pro Val Asn Met Thr Asp Cys Arg Leu Thr Ser Gly Lys Tyr Pro
100 105 110
Gln Cys Arg Tyr Ser Ala Ala Ala Gln Tyr Lys Phe Phe Ile Val Ala
115 120 125
Cys Asp Pro Pro Gln Lys Ser Asp Pro Pro Tyr Lys Leu Val Pro Val

```
    130                 135                 140
His Leu Asp Ser Ile Leu
145                 150

      <210> 101
      <211> 24
      <212> PRT
      <213> Homo sapiens

      <400> 101
Met Thr Pro Ser Pro Leu Leu Leu Leu Leu Leu Pro Pro Leu Leu Leu
 1               5                  10                  15
Gly Ala Phe Pro Pro Ala Ala Ala
            20

      <210> 102
      <211> 480
      <212> PRT
      <213> Homo sapiens

      <400> 102
Ala Arg Gly Pro Pro Lys Met Ala Asp Lys Val Val Pro Arg Gln Val
 1               5                  10                  15
Ala Arg Leu Gly Arg Thr Val Arg Leu Gln Cys Pro Val Glu Gly Asp
            20                  25                  30
Pro Pro Pro Leu Thr Met Trp Thr Lys Asp Gly Arg Thr Ile His Ser
        35                  40                  45
Gly Trp Ser Arg Phe Arg Val Leu Pro Gln Gly Leu Lys Val Lys Gln
    50                  55                  60
Val Glu Arg Glu Asp Ala Gly Val Tyr Val Cys Lys Ala Thr Asn Gly
65                  70                  75                  80
Phe Gly Ser Leu Ser Val Asn Tyr Thr Leu Val Val Leu Asp Asp Ile
                85                  90                  95
Ser Pro Gly Lys Glu Ser Leu Gly Pro Asp Ser Ser Ser Gly Gly Gln
            100                 105                 110
Glu Asp Pro Ala Ser Gln Gln Trp Ala Arg Pro Arg Phe Thr Gln Pro
        115                 120                 125
Ser Lys Met Arg Arg Arg Val Ile Ala Arg Pro Val Gly Ser Ser Val
    130                 135                 140
Arg Leu Lys Cys Val Ala Ser Gly His Pro Arg Pro Asp Ile Thr Trp
145                 150                 155                 160
Met Lys Asp Asp Gln Ala Leu Thr Arg Pro Glu Ala Ala Glu Pro Arg
                165                 170                 175
Lys Lys Lys Trp Thr Leu Ser Leu Lys Asn Leu Arg Pro Glu Asp Ser
            180                 185                 190
Gly Lys Tyr Thr Cys Arg Val Ser Asn Arg Ala Gly Ala Ile Asn Ala
        195                 200                 205
Thr Tyr Lys Val Asp Val Ile Gln Arg Thr Arg Ser Lys Pro Val Leu
    210                 215                 220
Thr Gly Thr His Pro Val Asn Thr Thr Val Asp Phe Gly Gly Thr Thr
225                 230                 235                 240
Ser Phe Gln Cys Lys Val Arg Ser Asp Val Lys Pro Val Ile Gln Trp
                245                 250                 255
Leu Lys Arg Val Glu Tyr Gly Ala Glu Gly Arg His Asn Ser Thr Ile
            260                 265                 270
Asp Val Gly Gly Gln Lys Phe Val Val Leu Pro Thr Gly Asp Val Trp
        275                 280                 285
Ser Arg Pro Asp Gly Ser Tyr Leu Asn Lys Leu Leu Ile Thr Arg Ala
```

```
    290                 295                 300
Arg Gln Asp Asp Ala Gly Met Tyr Ile Cys Leu Gly Ala Asn Thr Met
305                 310                 315                 320
Gly Tyr Ser Phe Arg Ser Ala Phe Leu Thr Val Leu Pro Asp Pro Lys
                325                 330                 335
Pro Pro Gly Pro Pro Val Ala Ser Ser Ser Ser Ala Thr Ser Leu Pro
            340                 345                 350
Trp Pro Val Val Ile Gly Ile Pro Ala Gly Ala Val Phe Ile Leu Gly
        355                 360                 365
Thr Leu Leu Leu Trp Leu Cys Gln Ala Gln Lys Lys Pro Cys Thr Pro
    370                 375                 380
Ala Pro Ala Pro Pro Leu Pro Gly His Arg Pro Pro Gly Thr Ala Arg
385                 390                 395                 400
Asp Arg Ser Gly Asp Lys Asp Leu Pro Ser Leu Ala Ala Leu Ser Ala
                405                 410                 415
Gly Pro Gly Val Gly Leu Cys Glu Glu His Gly Ser Pro Ala Ala Pro
            420                 425                 430
Gln His Leu Leu Gly Pro Gly Pro Val Ala Gly Pro Lys Leu Tyr Pro
        435                 440                 445
Lys Leu Tyr Thr Asp Ile His Thr His Thr His Thr His Ser His Thr
    450                 455                 460
His Ser His Val Glu Gly Lys Val His Gln His Ile His Tyr Gln Cys
465                 470                 475                 480

   <210> 103
   <211> 350
   <212> PRT
   <213> Homo sapiens

   <400> 103
Ala Arg Gly Pro Pro Lys Met Ala Asp Lys Val Val Pro Arg Gln Val
 1               5                  10                  15
Ala Arg Leu Gly Arg Thr Val Arg Leu Gln Cys Pro Val Glu Gly Asp
            20                  25                  30
Pro Pro Pro Leu Thr Met Trp Thr Lys Asp Gly Arg Thr Ile His Ser
        35                  40                  45
Gly Trp Ser Arg Phe Arg Val Leu Pro Gln Gly Leu Lys Val Lys Gln
    50                  55                  60
Val Glu Arg Glu Asp Ala Gly Val Tyr Val Cys Lys Ala Thr Asn Gly
65                  70                  75                  80
Phe Gly Ser Leu Ser Val Asn Tyr Thr Leu Val Val Leu Asp Asp Ile
                85                  90                  95
Ser Pro Gly Lys Glu Ser Leu Gly Pro Asp Ser Ser Ser Gly Gly Gln
            100                 105                 110
Glu Asp Pro Ala Ser Gln Gln Trp Ala Arg Pro Arg Phe Thr Gln Pro
        115                 120                 125
Ser Lys Met Arg Arg Arg Val Ile Ala Arg Pro Val Gly Ser Ser Val
    130                 135                 140
Arg Leu Lys Cys Val Ala Ser Gly His Pro Arg Pro Asp Ile Thr Trp
145                 150                 155                 160
Met Lys Asp Asp Gln Ala Leu Thr Arg Pro Glu Ala Ala Glu Pro Arg
                165                 170                 175
Lys Lys Lys Trp Thr Leu Ser Leu Lys Asn Leu Arg Pro Glu Asp Ser
            180                 185                 190
Gly Lys Tyr Thr Cys Arg Val Ser Asn Arg Ala Gly Ala Ile Asn Ala
        195                 200                 205
Thr Tyr Lys Val Asp Val Ile Gln Arg Thr Arg Ser Lys Pro Val Leu
    210                 215                 220
```

Thr Gly Thr His Pro Val Asn Thr Thr Val Asp Phe Gly Gly Thr Thr
225 230 235 240
Ser Phe Gln Cys Lys Val Arg Ser Asp Val Lys Pro Val Ile Gln Trp
245 250 255
Leu Lys Arg Val Glu Tyr Gly Ala Glu Gly Arg His Asn Ser Thr Ile
260 265 270
Asp Val Gly Gly Gln Lys Phe Val Val Leu Pro Thr Gly Asp Val Trp
275 280 285
Ser Arg Pro Asp Gly Ser Tyr Leu Asn Lys Leu Leu Ile Thr Arg Ala
290 295 300
Arg Gln Asp Asp Ala Gly Met Tyr Ile Cys Leu Gly Ala Asn Thr Met
305 310 315 320
Gly Tyr Ser Phe Arg Ser Ala Phe Leu Thr Val Leu Pro Asp Pro Lys
325 330 335
Pro Pro Gly Pro Pro Val Ala Ser Ser Ser Ser Ala Thr Ser
340 345 350

<210> 104
<211> 24
<212> PRT
<213> Homo sapiens

<400> 104
Leu Pro Trp Pro Val Val Ile Gly Ile Pro Ala Gly Ala Val Phe Ile
1 5 10 15
Leu Gly Thr Leu Leu Leu Trp Leu
20

<210> 105
<211> 106
<212> PRT
<213> Homo sapiens

<400> 105
Cys Gln Ala Gln Lys Lys Pro Cys Thr Pro Ala Pro Ala Pro Pro Leu
1 5 10 15
Pro Gly His Arg Pro Pro Gly Thr Ala Arg Asp Arg Ser Gly Asp Lys
20 25 30
Asp Leu Pro Ser Leu Ala Ala Leu Ser Ala Gly Pro Gly Val Gly Leu
35 40 45
Cys Glu Glu His Gly Ser Pro Ala Ala Pro Gln His Leu Leu Gly Pro
50 55 60
Gly Pro Val Ala Gly Pro Lys Leu Tyr Pro Lys Leu Tyr Thr Asp Ile
65 70 75 80
His Thr His Thr His Thr His Ser His Thr His Ser His Val Glu Gly
85 90 95
Lys Val His Gln His Ile His Tyr Gln Cys
100 105

<210> 106
<211> 208
<212> PRT
<213> Mus musculus

<400> 106
Arg Val Arg Pro Thr Gly Asp Val Trp Ser Arg Pro Asp Gly Ser Tyr
1 5 10 15
Leu Asn Lys Leu Leu Ile Ser Arg Ala Arg Gln Asp Asp Ala Gly Met

```
            20                  25                  30
Tyr Ile Cys Leu Gly Ala Asn Thr Met Gly Tyr Ser Phe Arg Ser Ala
        35                  40                  45
Phe Leu Thr Val Leu Pro Asp Pro Lys Pro Pro Gly Pro Pro Met Ala
    50                  55                  60
Ser Ser Ser Ser Ser Thr Ser Leu Pro Trp Pro Val Val Ile Gly Ile
65                  70                  75                  80
Pro Ala Gly Ala Val Phe Ile Leu Gly Thr Val Leu Leu Trp Leu Cys
                85                  90                  95
Gln Thr Lys Lys Lys Pro Cys Ala Pro Ala Ser Thr Leu Pro Val Pro
            100                 105                 110
Gly His Arg Pro Pro Gly Thr Ser Arg Glu Arg Ser Gly Asp Lys Asp
        115                 120                 125
Leu Pro Ser Leu Ala Val Gly Ile Cys Glu Glu His Gly Ser Ala Met
    130                 135                 140
Ala Pro Gln His Ile Leu Ala Ser Gly Ser Thr Ala Gly Pro Lys Leu
145                 150                 155                 160
Tyr Pro Lys Leu Tyr Thr Asp Val His Thr His Thr His Thr His Thr
                165                 170                 175
Cys Thr His Thr Leu Ser Cys Trp Arg Ala Arg Phe Ile Asn Thr Ser
            180                 185                 190
Met Ser Thr Ile Ser Ala Lys Tyr Ser Glu Ser Pro Ser Thr Val Ser
        195                 200                 205
```

```
<210> 107
<211> 73
<212> PRT
<213> Mus musculus

<400> 107
Arg Val Arg Pro Thr Gly Asp Val Trp Ser Arg Pro Asp Gly Ser Tyr
 1               5                  10                  15
Leu Asn Lys Leu Leu Ile Ser Arg Ala Arg Gln Asp Asp Ala Gly Met
            20                  25                  30
Tyr Ile Cys Leu Gly Ala Asn Thr Met Gly Tyr Ser Phe Arg Ser Ala
        35                  40                  45
Phe Leu Thr Val Leu Pro Asp Pro Lys Pro Pro Gly Pro Pro Met Ala
    50                  55                  60
Ser Ser Ser Ser Ser Thr Ser Leu Pro
65                  70
```

```
<210> 108
<211> 23
<212> PRT
<213> Mus musculus

<400> 108
Trp Pro Val Val Ile Gly Ile Pro Ala Gly Ala Val Phe Ile Leu Gly
 1               5                  10                  15
Thr Val Leu Leu Trp Leu Cys
            20
```

```
<210> 109
<211> 112
<212> PRT
<213> Mus musculus

<400> 109
```

```
Gln Thr Lys Lys Lys Pro Cys Ala Pro Ala Ser Thr Leu Pro Val Pro
 1               5                  10                  15
Gly His Arg Pro Pro Gly Thr Ser Arg Glu Arg Ser Gly Asp Lys Asp
            20                  25                  30
Leu Pro Ser Leu Ala Val Gly Ile Cys Glu Glu His Gly Ser Ala Met
        35                  40                  45
Ala Pro Gln His Ile Leu Ala Ser Gly Ser Thr Ala Gly Pro Lys Leu
    50                  55                  60
Tyr Pro Lys Leu Tyr Thr Asp Val His Thr His Thr His Thr His Thr
65                  70                  75                  80
Cys Thr His Thr Leu Ser Cys Trp Arg Ala Arg Phe Ile Asn Thr Ser
                85                  90                  95
Met Ser Thr Ile Ser Ala Lys Tyr Ser Glu Ser Pro Ser Thr Val Ser
            100                 105                 110

      <210> 110
      <211> 35
      <212> PRT
      <213> Homo sapiens

      <400> 110
Met Pro Gly Pro Arg Val Trp Gly Lys Tyr Leu Trp Arg Ser Pro His
 1               5                  10                  15
Ser Lys Gly Cys Pro Gly Ala Met Trp Trp Leu Leu Leu Trp Gly Val
            20                  25                  30
Leu Gln Ala
        35

      <210> 111
      <211> 103
      <212> PRT
      <213> Homo sapiens

      <400> 111
Cys Pro Thr Arg Gly Ser Val Leu Leu Ala Gln Glu Leu Pro Gln Gln
 1               5                  10                  15
Leu Thr Ser Pro Gly Tyr Pro Glu Pro Tyr Gly Lys Gly Gln Glu Ser
            20                  25                  30
Ser Thr Asp Ile Lys Ala Pro Glu Gly Phe Ala Val Arg Leu Val Phe
        35                  40                  45
Gln Asp Phe Asp Leu Glu Pro Ser Gln Asp Cys Ala Gly Asp Ser Val
    50                  55                  60
Thr Val Ser Trp Gly Trp Gly Gly Ser Arg Gln Asp Cys Gly Gln Gly
65                  70                  75                  80
Asp Ser Arg Gly Cys Gly Lys Trp Arg Cys Pro Glu Ser Pro Ile Trp
                85                  90                  95
Arg Arg Asp Glu Phe Ser Met
            100

      <210> 112
      <211> 20
      <212> PRT
      <213> Homo sapiens

      <400> 112
Met Ser Pro Pro Leu Cys Pro Leu Leu Leu Leu Ala Val Gly Leu Arg
 1               5                  10                  15
Leu Ala Gly Thr
```

20

<210> 113
<211> 1030
<212> PRT
<213> Homo sapiens

<400> 113
Leu Asn Pro Ser Asp Pro Asn Thr Cys Ser Phe Trp Glu Ser Phe Thr
1 5 10 15
Thr Thr Thr Lys Glu Ser His Ser Arg Pro Phe Ser Leu Leu Pro Ser
20 25 30
Glu Pro Cys Glu Arg Pro Trp Glu Gly Pro His Thr Cys Pro Ser Pro
35 40 45
Gln Thr Gln Arg Lys Leu Leu Ala Ser Arg Asp Ser Phe Cys Met Val
50 55 60
Cys Val Gly Ala Gly Val Gln Trp Arg Asp Arg Ser Ala Leu Gln Pro
65 70 75 80
Gln Thr Gly Asn Ala Leu Ser Met Arg Pro Gln Pro Arg Val Leu Ser
85 90 95
Gly Ala Pro Ser Leu Ala Ser Pro Gly His Thr Val Val Val Lys Thr
100 105 110
Asp His Arg Gln Arg Leu Gln Cys Cys His Gly Phe Tyr Glu Ser Arg
115 120 125
Gly Phe Cys Val Pro Leu Cys Ala Gln Glu Cys Val His Gly Arg Cys
130 135 140
Val Ala Pro Asn Gln Cys Gln Cys Val Pro Gly Trp Arg Gly Asp Asp
145 150 155 160
Cys Ser Ser Ala Pro Asn Cys Leu Gln Pro Cys Thr Pro Gly Tyr Tyr
165 170 175
Gly Pro Ala Cys Gln Phe Arg Cys Gln Cys His Gly Ala Pro Cys Asp
180 185 190
Pro Gln Thr Gly Ala Cys Phe Cys Pro Ala Glu Arg Thr Gly Pro Ser
195 200 205
Cys Asp Val Ser Cys Ser Gln Gly Thr Ser Gly Phe Phe Cys Pro Ser
210 215 220
Thr His Pro Cys Gln Asn Gly Gly Val Phe Gln Thr Pro Gln Gly Ser
225 230 235 240
Cys Ser Cys Pro Pro Gly Trp Met Gly Thr Ile Cys Ser Leu Pro Cys
245 250 255
Pro Glu Gly Phe His Gly Pro Asn Cys Ser Gln Glu Cys Arg Cys His
260 265 270
Asn Gly Gly Leu Cys Asp Arg Phe Thr Gly Gln Cys Arg Cys Ala Pro
275 280 285
Gly Tyr Thr Gly Asp Arg Cys Arg Glu Glu Cys Pro Val Gly Arg Phe
290 295 300
Gly Gln Asp Cys Ala Glu Thr Cys Asp Cys Ala Pro Asp Ala Arg Cys
305 310 315 320
Phe Pro Ala Asn Gly Ala Cys Leu Cys Glu His Gly Phe Thr Gly Asp
325 330 335
Arg Cys Thr Asp Arg Leu Cys Pro Asp Gly Phe Tyr Gly Leu Ser Cys
340 345 350
Gln Ala Pro Cys Thr Cys Asp Arg Glu His Ser Leu Ser Cys His Pro
355 360 365
Met Asn Gly Glu Cys Ser Cys Leu Pro Gly Trp Ala Gly Leu His Cys
370 375 380
Asn Glu Ser Cys Pro Gln Asp Thr His Gly Pro Gly Cys Gln Glu His
385 390 395 400

```
Cys Leu Cys Leu His Gly Gly Val Cys Gln Ala Thr Ser Gly Leu Cys
                405                 410                 415
Gln Cys Ala Pro Gly Tyr Thr Gly Pro His Cys Ala Ser Leu Cys Pro
            420                 425                 430
Pro Asp Thr Tyr Gly Val Asn Cys Ser Ala Arg Cys Ser Cys Glu Asn
        435                 440                 445
Ala Ile Ala Cys Ser Pro Ile Asp Gly Glu Cys Val Cys Lys Glu Gly
    450                 455                 460
Trp Gln Arg Gly Asn Cys Ser Val Pro Cys Pro Pro Gly Thr Trp Gly
465                 470                 475                 480
Phe Ser Cys Asn Ala Ser Cys Gln Cys Ala His Glu Ala Val Cys Ser
                485                 490                 495
Pro Gln Thr Gly Ala Cys Thr Cys Thr Pro Gly Trp His Gly Ala His
            500                 505                 510
Cys Gln Leu Pro Cys Pro Lys Gly Gln Phe Gly Glu Gly Cys Ala Ser
        515                 520                 525
Arg Cys Asp Cys Asp His Ser Asp Gly Cys Asp Pro Val His Gly Arg
    530                 535                 540
Cys Gln Cys Gln Ala Gly Trp Met Gly Ala Arg Cys His Leu Ser Cys
545                 550                 555                 560
Pro Glu Gly Leu Trp Gly Val Asn Cys Ser Asn Thr Cys Thr Cys Lys
                565                 570                 575
Asn Gly Gly Thr Cys Leu Pro Glu Asn Gly Asn Cys Val Cys Ala Pro
            580                 585                 590
Gly Phe Arg Gly Pro Ser Cys Gln Arg Ser Cys Gln Pro Gly Arg Tyr
        595                 600                 605
Gly Lys Arg Cys Val Pro Cys Lys Cys Ala Asn His Ser Phe Cys His
    610                 615                 620
Pro Ser Asn Gly Thr Cys Tyr Cys Leu Ala Gly Trp Thr Gly Pro Asp
625                 630                 635                 640
Cys Ser Gln Pro Cys Pro Pro Gly His Trp Gly Glu Asn Cys Ala Gln
                645                 650                 655
Thr Cys Gln Cys His His Gly Gly Thr Cys His Pro Gln Asp Gly Ser
            660                 665                 670
Cys Ile Cys Pro Leu Gly Trp Thr Gly His His Cys Leu Glu Gly Cys
        675                 680                 685
Pro Leu Gly Thr Phe Gly Ala Asn Cys Ser Gln Pro Cys Gln Cys Gly
    690                 695                 700
Pro Gly Glu Lys Cys His Pro Glu Thr Gly Ala Cys Val Cys Pro Pro
705                 710                 715                 720
Gly His Ser Gly Ala Pro Cys Arg Ile Gly Ile Gln Glu Pro Phe Thr
                725                 730                 735
Val Met Pro Thr Thr Pro Val Ala Tyr Asn Ser Leu Gly Ala Val Ile
            740                 745                 750
Gly Ile Ala Val Leu Gly Ser Leu Val Val Ala Leu Val Ala Leu Phe
        755                 760                 765
Ile Gly Tyr Arg His Trp Gln Lys Gly Lys Glu His His His Leu Ala
    770                 775                 780
Val Ala Tyr Ser Ser Gly Arg Leu Asp Gly Ser Glu Tyr Val Met Pro
785                 790                 795                 800
Asp Val Pro Pro Ser Tyr Ser His Tyr Tyr Ser Asn Pro Ser Tyr His
                805                 810                 815
Thr Leu Ser Gln Cys Ser Pro Asn Pro Pro Pro Pro Asn Lys Val Pro
            820                 825                 830
Gly Pro Leu Phe Ala Ser Leu Gln Asn Pro Glu Arg Pro Gly Gly Ala
        835                 840                 845
Gln Gly His Asp Asn His Thr Thr Leu Pro Ala Asp Trp Lys His Arg
    850                 855                 860
```

```
Arg Glu Pro Pro Pro Gly Pro Leu Asp Arg Gly Ser Ser Arg Leu Asp
865                 870                 875                 880
Arg Ser Tyr Ser Tyr Ser Tyr Ser Asn Gly Pro Gly Pro Phe Tyr Asp
                885                 890                 895
Lys Gly Leu Ile Ser Glu Glu Glu Leu Gly Ala Ser Val Ala Ser Leu
            900                 905                 910
Ser Ser Glu Asn Pro Tyr Ala Thr Ile Arg Asp Leu Pro Ser Leu Pro
        915                 920                 925
Gly Gly Pro Arg Glu Ser Ser Tyr Met Glu Met Lys Gly Pro Pro Ser
    930                 935                 940
Gly Ser Ala Pro Arg Gln Pro Pro Gln Phe Trp Asp Ser Gln Arg Arg
945                 950                 955                 960
Arg Gln Pro Gln Pro Gln Arg Asp Ser Gly Thr Tyr Glu Gln Pro Ser
                965                 970                 975
Pro Leu Ile His Asp Arg Asp Ser Val Gly Ser Gln Pro Pro Leu Pro
            980                 985                 990
Pro Gly Leu Pro Pro Gly His Tyr Asp Ser Pro Lys Asn Ser His Ile
        995                 1000                1005
Pro Gly His Tyr Asp Leu Pro Pro Val Arg His Pro Pro Ser Pro Pro
    1010                1015                1020
Leu Arg Arg Gln Asp Arg
1025                1030

      &lt;210&gt; 114
      &lt;211&gt; 747
      &lt;212&gt; PRT
      &lt;213&gt; Homo sapiens

      &lt;400&gt; 114
Leu Asn Pro Ser Asp Pro Asn Thr Cys Ser Phe Trp Glu Ser Phe Thr
 1               5                  10                  15
Thr Thr Thr Lys Glu Ser His Ser Arg Pro Phe Ser Leu Leu Pro Ser
            20                  25                  30
Glu Pro Cys Glu Arg Pro Trp Glu Gly Pro His Thr Cys Pro Ser Pro
        35                  40                  45
Gln Thr Gln Arg Lys Leu Leu Ala Ser Arg Asp Ser Phe Cys Met Val
    50                  55                  60
Cys Val Gly Ala Gly Val Gln Trp Arg Asp Arg Ser Ala Leu Gln Pro
65                  70                  75                  80
Gln Thr Gly Asn Ala Leu Ser Met Arg Pro Gln Pro Arg Val Leu Ser
                85                  90                  95
Gly Ala Pro Ser Leu Ala Ser Pro Gly His Thr Val Val Val Lys Thr
            100                 105                 110
Asp His Arg Gln Arg Leu Gln Cys Cys His Gly Phe Tyr Glu Ser Arg
        115                 120                 125
Gly Phe Cys Val Pro Leu Cys Ala Gln Glu Cys Val His Gly Arg Cys
    130                 135                 140
Val Ala Pro Asn Gln Cys Gln Cys Val Pro Gly Trp Arg Gly Asp Asp
145                 150                 155                 160
Cys Ser Ser Ala Pro Asn Cys Leu Gln Pro Cys Thr Pro Gly Tyr Tyr
                165                 170                 175
Gly Pro Ala Cys Gln Phe Arg Cys Gln Cys His Gly Ala Pro Cys Asp
            180                 185                 190
Pro Gln Thr Gly Ala Cys Phe Cys Pro Ala Glu Arg Thr Gly Pro Ser
        195                 200                 205
Cys Asp Val Ser Cys Ser Gln Gly Thr Ser Gly Phe Phe Cys Pro Ser
    210                 215                 220
Thr His Pro Cys Gln Asn Gly Gly Val Phe Gln Thr Pro Gln Gly Ser
```

```
225                 230                 235                 240
Cys Ser Cys Pro Pro Gly Trp Met Gly Thr Ile Cys Ser Leu Pro Cys
                245                 250                 255
Pro Glu Gly Phe His Gly Pro Asn Cys Ser Gln Glu Cys Arg Cys His
            260                 265                 270
Asn Gly Gly Leu Cys Asp Arg Phe Thr Gly Gln Cys Arg Cys Ala Pro
        275                 280                 285
Gly Tyr Thr Gly Asp Arg Cys Arg Glu Glu Cys Pro Val Gly Arg Phe
    290                 295                 300
Gly Gln Asp Cys Ala Glu Thr Cys Asp Cys Ala Pro Asp Ala Arg Cys
305                 310                 315                 320
Phe Pro Ala Asn Gly Ala Cys Leu Cys Glu His Gly Phe Thr Gly Asp
                325                 330                 335
Arg Cys Thr Asp Arg Leu Cys Pro Asp Gly Phe Tyr Gly Leu Ser Cys
            340                 345                 350
Gln Ala Pro Cys Thr Cys Asp Arg Glu His Ser Leu Ser Cys His Pro
        355                 360                 365
Met Asn Gly Glu Cys Ser Cys Leu Pro Gly Trp Ala Gly Leu His Cys
    370                 375                 380
Asn Glu Ser Cys Pro Gln Asp Thr His Gly Pro Gly Cys Gln Glu His
385                 390                 395                 400
Cys Leu Cys Leu His Gly Gly Val Cys Gln Ala Thr Ser Gly Leu Cys
                405                 410                 415
Gln Cys Ala Pro Gly Tyr Thr Gly Pro His Cys Ala Ser Leu Cys Pro
            420                 425                 430
Pro Asp Thr Tyr Gly Val Asn Cys Ser Ala Arg Cys Ser Cys Glu Asn
        435                 440                 445
Ala Ile Ala Cys Ser Pro Ile Asp Gly Glu Cys Val Cys Lys Glu Gly
    450                 455                 460
Trp Gln Arg Gly Asn Cys Ser Val Pro Cys Pro Pro Gly Thr Trp Gly
465                 470                 475                 480
Phe Ser Cys Asn Ala Ser Cys Gln Cys Ala His Glu Ala Val Cys Ser
                485                 490                 495
Pro Gln Thr Gly Ala Cys Thr Cys Thr Pro Gly Trp His Gly Ala His
            500                 505                 510
Cys Gln Leu Pro Cys Pro Lys Gly Gln Phe Gly Glu Gly Cys Ala Ser
        515                 520                 525
Arg Cys Asp Cys Asp His Ser Asp Gly Cys Asp Pro Val His Gly Arg
    530                 535                 540
Cys Gln Cys Gln Ala Gly Trp Met Gly Ala Arg Cys His Leu Ser Cys
545                 550                 555                 560
Pro Glu Gly Leu Trp Gly Val Asn Cys Ser Asn Thr Cys Thr Cys Lys
                565                 570                 575
Asn Gly Gly Thr Cys Leu Pro Glu Asn Gly Asn Cys Val Cys Ala Pro
            580                 585                 590
Gly Phe Arg Gly Pro Ser Cys Gln Arg Ser Cys Gln Pro Gly Arg Tyr
        595                 600                 605
Gly Lys Arg Cys Val Pro Cys Lys Cys Ala Asn His Ser Phe Cys His
    610                 615                 620
Pro Ser Asn Gly Thr Cys Tyr Cys Leu Ala Gly Trp Thr Gly Pro Asp
625                 630                 635                 640
Cys Ser Gln Pro Cys Pro Pro Gly His Trp Gly Glu Asn Cys Ala Gln
                645                 650                 655
Thr Cys Gln Cys His His Gly Gly Thr Cys His Pro Gln Asp Gly Ser
            660                 665                 670
Cys Ile Cys Pro Leu Gly Trp Thr Gly His His Cys Leu Glu Gly Cys
        675                 680                 685
Pro Leu Gly Thr Phe Gly Ala Asn Cys Ser Gln Pro Cys Gln Cys Gly
```

```
    690                 695                 700
Pro Gly Glu Lys Cys His Pro Glu Thr Gly Ala Cys Val Cys Pro Pro
705                 710                 715                 720
Gly His Ser Gly Ala Pro Cys Arg Ile Gly Ile Gln Glu Pro Phe Thr
                725                 730                 735
Val Met Pro Thr Thr Pro Val Ala Tyr Asn Ser
            740                 745

      <210> 115
      <211> 24
      <212> PRT
      <213> Homo sapiens

      <400> 115
Leu Gly Ala Val Ile Gly Ile Ala Val Leu Gly Ser Leu Val Val Ala
 1               5                  10                  15
Leu Val Ala Leu Phe Ile Gly Tyr
            20

      <210> 116
      <211> 259
      <212> PRT
      <213> Homo sapiens

      <400> 116
Arg His Trp Gln Lys Gly Lys Glu His His His Leu Ala Val Ala Tyr
 1               5                  10                  15
Ser Ser Gly Arg Leu Asp Gly Ser Glu Tyr Val Met Pro Asp Val Pro
            20                  25                  30
Pro Ser Tyr Ser His Tyr Tyr Ser Asn Pro Ser Tyr His Thr Leu Ser
        35                  40                  45
Gln Cys Ser Pro Asn Pro Pro Pro Pro Asn Lys Val Pro Gly Pro Leu
    50                  55                  60
Phe Ala Ser Leu Gln Asn Pro Glu Arg Pro Gly Gly Ala Gln Gly His
65                  70                  75                  80
Asp Asn His Thr Thr Leu Pro Ala Asp Trp Lys His Arg Arg Glu Pro
                85                  90                  95
Pro Pro Gly Pro Leu Asp Arg Gly Ser Ser Arg Leu Asp Arg Ser Tyr
            100                 105                 110
Ser Tyr Ser Tyr Ser Asn Gly Pro Gly Pro Phe Tyr Asp Lys Gly Leu
        115                 120                 125
Ile Ser Glu Glu Glu Leu Gly Ala Ser Val Ala Ser Leu Ser Ser Glu
    130                 135                 140
Asn Pro Tyr Ala Thr Ile Arg Asp Leu Pro Ser Leu Pro Gly Gly Pro
145                 150                 155                 160
Arg Glu Ser Ser Tyr Met Glu Met Lys Gly Pro Pro Ser Gly Ser Ala
                165                 170                 175
Pro Arg Gln Pro Pro Gln Phe Trp Asp Ser Gln Arg Arg Arg Gln Pro
            180                 185                 190
Gln Pro Gln Arg Asp Ser Gly Thr Tyr Glu Gln Pro Ser Pro Leu Ile
        195                 200                 205
His Asp Arg Asp Ser Val Gly Ser Gln Pro Pro Leu Pro Pro Gly Leu
    210                 215                 220
Pro Pro Gly His Tyr Asp Ser Pro Lys Asn Ser His Ile Pro Gly His
225                 230                 235                 240
Tyr Asp Leu Pro Pro Val Arg His Pro Pro Ser Pro Pro Leu Arg Arg
                245                 250                 255
Gln Asp Arg
```

<210> 117
<211> 497
<212> PRT
<213> Mus msuculus

<400> 117
Ser Thr His Ala Ser Gly Asp Pro Val His Gly Gln Cys Arg Cys Gln
1 5 10 15
Ala Gly Trp Met Gly Thr Arg Cys His Leu Pro Cys Pro Glu Gly Phe
20 25 30
Trp Gly Ala Asn Cys Ser Asn Thr Cys Thr Cys Lys Asn Gly Gly Thr
35 40 45
Cys Val Ser Glu Asn Gly Asn Cys Val Cys Ala Pro Gly Phe Arg Gly
50 55 60
Pro Ser Cys Gln Arg Pro Cys Pro Pro Gly Arg Tyr Gly Lys Arg Cys
65 70 75 80
Val Gln Cys Lys Cys Asn Asn Asn His Ser Ser Cys His Pro Ser Asp
85 90 95
Gly Thr Cys Ser Cys Leu Ala Gly Trp Thr Gly Pro Asp Cys Ser Glu
100 105 110
Ala Cys Pro Pro Gly His Trp Gly Leu Lys Cys Ser Gln Leu Cys Gln
115 120 125
Cys His His Gly Gly Thr Cys His Pro Gln Asp Gly Ser Cys Ile Cys
130 135 140
Thr Pro Gly Trp Thr Gly Pro Asn Cys Leu Glu Gly Cys Pro Pro Arg
145 150 155 160
Met Phe Gly Val Asn Cys Ser Gln Leu Cys Gln Cys Asp Leu Gly Glu
165 170 175
Met Cys His Pro Glu Thr Gly Ala Cys Val Cys Pro Pro Gly His Ser
180 185 190
Gly Ala Asp Cys Lys Met Gly Ser Gln Glu Ser Phe Thr Ile Met Pro
195 200 205
Thr Ser Pro Val Thr His Asn Ser Leu Gly Ala Val Ile Gly Ile Ala
210 215 220
Val Leu Gly Thr Leu Val Val Ala Leu Ile Ala Leu Phe Ile Gly Tyr
225 230 235 240
Arg Gln Trp Gln Lys Gly Lys Glu His Glu His Leu Ala Val Ala Tyr
245 250 255
Ser Thr Gly Arg Leu Asp Gly Ser Asp Tyr Val Met Pro Asp Val Ser
260 265 270
Pro Ser Tyr Ser His Tyr Tyr Ser Asn Pro Ser Tyr His Thr Leu Ser
275 280 285
Gln Cys Ser Pro Asn Pro Pro Pro Pro Asn Lys Val Pro Gly Ser Gln
290 295 300
Leu Phe Val Ser Ser Gln Ala Pro Glu Arg Pro Ser Arg Ala His Gly
305 310 315 320
Arg Glu Asn His Thr Thr Leu Pro Ala Asp Trp Lys His Arg Arg Glu
325 330 335
Pro His Asp Arg Gly Ala Ser His Leu Asp Arg Ser Tyr Ser Cys Ser
340 345 350
Tyr Ser His Arg Asn Gly Pro Gly Pro Phe Cys His Lys Gly Pro Ile
355 360 365
Ser Glu Glu Gly Leu Gly Ala Ser Val Met Ser Leu Ser Ser Glu Asn
370 375 380
Pro Tyr Ala Thr Ile Arg Asp Leu Pro Ser Leu Pro Gly Glu Pro Arg
385 390 395 400

```
Glu Ser Gly Tyr Val Glu Met Lys Gly Pro Pro Ser Val Ser Pro Pro
                405                 410                 415
Arg Gln Ser Leu His Leu Arg Asp Arg Gln Gln Arg Gln Leu Gln Pro
            420                 425                 430
Gln Arg Asp Ser Gly Thr Tyr Glu Gln Pro Ser Pro Leu Ser His Asn
        435                 440                 445
Glu Glu Ser Leu Gly Ser Thr Pro Pro Leu Pro Pro Gly Leu Pro Pro
    450                 455                 460
Gly His Tyr Asp Ser Pro Lys Asn Ser His Ile Pro Gly His Tyr Asp
465                 470                 475                 480
Leu Pro Pro Val Arg His Pro Pro Ser Pro Pro Ser Arg Arg Gln Asp
                485                 490                 495
Arg


<210> 118
<211> 216
<212> PRT
<213> Mus musculus

<400> 118
Ser Thr His Ala Ser Gly Asp Pro Val His Gly Gln Cys Arg Cys Gln
 1               5                  10                  15
Ala Gly Trp Met Gly Thr Arg Cys His Leu Pro Cys Pro Glu Gly Phe
            20                  25                  30
Trp Gly Ala Asn Cys Ser Asn Thr Cys Thr Cys Lys Asn Gly Gly Thr
        35                  40                  45
Cys Val Ser Glu Asn Gly Asn Cys Val Cys Ala Pro Gly Phe Arg Gly
    50                  55                  60
Pro Ser Cys Gln Arg Pro Cys Pro Pro Gly Arg Tyr Gly Lys Arg Cys
65                  70                  75                  80
Val Gln Cys Lys Cys Asn Asn Asn His Ser Ser Cys His Pro Ser Asp
                85                  90                  95
Gly Thr Cys Ser Cys Leu Ala Gly Trp Thr Gly Pro Asp Cys Ser Glu
            100                 105                 110
Ala Cys Pro Pro Gly His Trp Gly Leu Lys Cys Ser Gln Leu Cys Gln
        115                 120                 125
Cys His His Gly Gly Thr Cys His Pro Gln Asp Gly Ser Cys Ile Cys
    130                 135                 140
Thr Pro Gly Trp Thr Gly Pro Asn Cys Leu Glu Gly Cys Pro Pro Arg
145                 150                 155                 160
Met Phe Gly Val Asn Cys Ser Gln Leu Cys Gln Cys Asp Leu Gly Glu
                165                 170                 175
Met Cys His Pro Glu Thr Gly Ala Cys Val Cys Pro Pro Gly His Ser
            180                 185                 190
Gly Ala Asp Cys Lys Met Gly Ser Gln Glu Ser Phe Thr Ile Met Pro
        195                 200                 205
Thr Ser Pro Val Thr His Asn Ser
    210                 215

<210> 119
<211> 24
<212> PRT
<213> Mus musculus

<400> 119
Leu Gly Ala Val Ile Gly Ile Ala Val Leu Gly Thr Leu Val Val Ala
 1               5                  10                  15
```

Leu Ile Ala Leu Phe Ile Gly Tyr
20

<210> 120
<211> 257
<212> PRT
<213> Mus musculus

<400> 120
Arg Gln Trp Gln Lys Gly Lys Glu His Glu His Leu Ala Val Ala Tyr
1 5 10 15
Ser Thr Gly Arg Leu Asp Gly Ser Asp Tyr Val Met Pro Asp Val Ser
20 25 30
Pro Ser Tyr Ser His Tyr Tyr Ser Asn Pro Ser Tyr His Thr Leu Ser
35 40 45
Gln Cys Ser Pro Asn Pro Pro Pro Pro Asn Lys Val Pro Gly Ser Gln
50 55 60
Leu Phe Val Ser Ser Gln Ala Pro Glu Arg Pro Ser Arg Ala His Gly
65 70 75 80
Arg Glu Asn His Thr Thr Leu Pro Ala Asp Trp Lys His Arg Arg Glu
85 90 95
Pro His Asp Arg Gly Ala Ser His Leu Asp Arg Ser Tyr Ser Cys Ser
100 105 110
Tyr Ser His Arg Asn Gly Pro Gly Pro Phe Cys His Lys Gly Pro Ile
115 120 125
Ser Glu Glu Gly Leu Gly Ala Ser Val Met Ser Leu Ser Ser Glu Asn
130 135 140
Pro Tyr Ala Thr Ile Arg Asp Leu Pro Ser Leu Pro Gly Glu Pro Arg
145 150 155 160
Glu Ser Gly Tyr Val Glu Met Lys Gly Pro Pro Ser Val Ser Pro Pro
165 170 175
Arg Gln Ser Leu His Leu Arg Asp Arg Gln Gln Arg Gln Leu Gln Pro
180 185 190
Gln Arg Asp Ser Gly Thr Tyr Glu Gln Pro Ser Pro Leu Ser His Asn
195 200 205
Glu Glu Ser Leu Gly Ser Thr Pro Pro Leu Pro Pro Gly Leu Pro Pro
210 215 220
Gly His Tyr Asp Ser Pro Lys Asn Ser His Ile Pro Gly His Tyr Asp
225 230 235 240
Leu Pro Pro Val Arg His Pro Pro Ser Pro Pro Ser Arg Arg Gln Asp
245 250 255
Arg

<210> 121
<211> 636
<212> PRT
<213> Rattus sp.

<400> 121
Met Gly Val Ile Cys Ser Leu Pro Cys Pro Glu Gly Phe His Gly Pro
1 5 10 15
Asn Cys Thr Gln Glu Cys Arg Cys His Asn Gly Gly Leu Cys Asp Arg
20 25 30
Phe Thr Gly Gln Cys His Cys Ala Pro Gly Tyr Ile Gly Asp Arg Cys
35 40 45
Arg Glu Glu Cys Pro Val Gly Arg Phe Gly Gln Asp Cys Ala Glu Thr
50 55 60

```
Cys Asp Cys Ala Pro Gly Ala Arg Cys Phe Pro Ala Asn Gly Ala Cys
65                  70                  75                  80
Leu Cys Glu His Gly Phe Thr Gly Asp Arg Cys Thr Glu Arg Leu Cys
                85                  90                  95
Pro Asp Gly Arg Tyr Gly Leu Ser Cys Gln Asp Pro Cys Thr Cys Asp
            100                 105                 110
Pro Glu His Ser Leu Ser Cys His Pro Met His Gly Glu Cys Ser Cys
        115                 120                 125
Gln Pro Gly Trp Ala Gly Leu His Cys Asn Glu Ser Cys Pro Gln Asp
    130                 135                 140
Thr His Gly Ala Gly Cys Gln Glu His Cys Leu Cys Leu His Gly Gly
145                 150                 155                 160
Val Cys Leu Ala Asp Ser Gly Leu Cys Arg Cys Ala Pro Gly Tyr Thr
                165                 170                 175
Gly Pro His Cys Ala Asn Leu Cys Pro Pro Asn Thr Tyr Gly Ile Asn
            180                 185                 190
Cys Ser Ser His Cys Ser Cys Glu Asn Ala Ile Ala Cys Ser Pro Val
        195                 200                 205
Asp Gly Thr Cys Ile Cys Lys Glu Gly Trp Gln Arg Gly Asn Cys Ser
    210                 215                 220
Val Pro Cys Pro Pro Gly Thr Trp Gly Phe Ser Cys Asn Ala Ser Cys
225                 230                 235                 240
Gln Cys Ala His Glu Gly Val Cys Ser Pro Gln Thr Gly Ala Cys Thr
                245                 250                 255
Cys Thr Pro Gly Trp Arg Gly Val His Cys Gln Leu Pro Cys Pro Lys
            260                 265                 270
Gly Gln Phe Gly Glu Gly Cys Ala Ser Val Cys Asp Cys Asp His Ser
        275                 280                 285
Asp Gly Cys Asp Pro Val His Gly His Cys Arg Cys Gln Ala Gly Trp
    290                 295                 300
Met Gly Thr Arg Cys His Leu Pro Cys Pro Glu Gly Phe Trp Gly Ala
305                 310                 315                 320
Asn Cys Ser Asn Ala Cys Thr Cys Lys Asn Gly Gly Thr Cys Val Pro
                325                 330                 335
Glu Asn Gly Asn Cys Val Cys Ala Pro Gly Phe Arg Gly Pro Ser Cys
            340                 345                 350
Gln Arg Pro Cys Pro Pro Gly Arg Tyr Gly Lys Arg Cys Val Pro Cys
        355                 360                 365
Lys Cys Asn Asn His Ser Ser Cys His Pro Ser Asp Gly Thr Cys Ser
    370                 375                 380
Cys Leu Ala Gly Trp Thr Gly Pro Asp Cys Ser Glu Ser Cys Pro Pro
385                 390                 395                 400
Gly His Trp Gly Leu Lys Cys Ser Gln Pro Cys Gln Cys His His Gly
                405                 410                 415
Ala Thr Cys His Pro Gln Asp Gly Ser Cys Val Cys Ile Pro Gly Trp
            420                 425                 430
Thr Gly Pro Asn Cys Ser Glu Gly Cys Pro Ser Arg Met Phe Gly Val
        435                 440                 445
Asn Cys Ser Gln Leu Cys Gln Cys Asp Pro Gly Glu Met Cys His Pro
    450                 455                 460
Glu Thr Gly Ala Cys Val Cys Pro Pro Gly His Ser Gly Ala His Cys
465                 470                 475                 480
Lys Val Gly Ser Gln Glu Ser Phe Thr Ile Met Pro Thr Ser Pro Val
                485                 490                 495
Ile His Asn Ser Leu Gly Ala Val Ile Gly Ile Ala Val Leu Gly Thr
            500                 505                 510
Leu Val Val Ala Leu Val Ala Leu Phe Ile Gly Tyr Arg His Trp Gln
        515                 520                 525
```

Lys Gly Lys Glu His Glu His Leu Ala Val Ala Tyr Ser Thr Gly Arg
    530   535   540
Leu Asp Gly Ser Asp Tyr Val Met Pro Asp Val Ser Pro Ser Tyr Ser
545   550   555   560
His Tyr Tyr Ser Asn Pro Ser Tyr His Thr Leu Ser Gln Cys Ser Pro
                565   570   575
Asn Pro Pro Pro Pro Asn Lys Ile Pro Gly Ser Gln Leu Phe Val Ser
            580   585   590
Ser Gln Ala Ser Glu Arg Pro Asn Arg Asn His Gly Arg Asp Asn His
        595   600   605
Ala Thr Leu Pro Ala Asp Trp Lys His Arg Arg Glu Ser His Asp Arg
    610   615   620
Ala Phe Leu Arg His Gln Pro Pro Gly Pro Lys Val
625   630   635

<210> 122
<211> 500
<212> PRT
<213> Rattus sp.

<400> 122
Met Gly Val Ile Cys Ser Leu Pro Cys Pro Glu Gly Phe His Gly Pro
1   5   10   15
Asn Cys Thr Gln Glu Cys Arg Cys His Asn Gly Gly Leu Cys Asp Arg
            20   25   30
Phe Thr Gly Gln Cys His Cys Ala Pro Gly Tyr Ile Gly Asp Arg Cys
        35   40   45
Arg Glu Glu Cys Pro Val Gly Arg Phe Gly Gln Asp Cys Ala Glu Thr
    50   55   60
Cys Asp Cys Ala Pro Gly Ala Arg Cys Phe Pro Ala Asn Gly Ala Cys
65   70   75   80
Leu Cys Glu His Gly Phe Thr Gly Asp Arg Cys Thr Glu Arg Leu Cys
                85   90   95
Pro Asp Gly Arg Tyr Gly Leu Ser Cys Gln Asp Pro Cys Thr Cys Asp
            100   105   110
Pro Glu His Ser Leu Ser Cys His Pro Met His Gly Glu Cys Ser Cys
        115   120   125
Gln Pro Gly Trp Ala Gly Leu His Cys Asn Glu Ser Cys Pro Gln Asp
    130   135   140
Thr His Gly Ala Gly Cys Gln Glu His Cys Leu Cys Leu His Gly Gly
145   150   155   160
Val Cys Leu Ala Asp Ser Gly Leu Cys Arg Cys Ala Pro Gly Tyr Thr
                165   170   175
Gly Pro His Cys Ala Asn Leu Cys Pro Pro Asn Thr Tyr Gly Ile Asn
            180   185   190
Cys Ser Ser His Cys Ser Cys Glu Asn Ala Ile Ala Cys Ser Pro Val
        195   200   205
Asp Gly Thr Cys Ile Cys Lys Glu Gly Trp Gln Arg Gly Asn Cys Ser
    210   215   220
Val Pro Cys Pro Pro Gly Thr Trp Gly Phe Ser Cys Asn Ala Ser Cys
225   230   235   240
Gln Cys Ala His Glu Gly Val Cys Ser Pro Gln Thr Gly Ala Cys Thr
                245   250   255
Cys Thr Pro Gly Trp Arg Gly Val His Cys Gln Leu Pro Cys Pro Lys
            260   265   270
Gly Gln Phe Gly Glu Gly Cys Ala Ser Val Cys Asp Cys Asp His Ser
        275   280   285
Asp Gly Cys Asp Pro Val His Gly His Cys Arg Cys Gln Ala Gly Trp

```
    290                     295                     300
Met Gly Thr Arg Cys His Leu Pro Cys Pro Glu Gly Phe Trp Gly Ala
305                 310                 315                 320
Asn Cys Ser Asn Ala Cys Thr Cys Lys Asn Gly Gly Thr Cys Val Pro
                325                 330                 335
Glu Asn Gly Asn Cys Val Cys Ala Pro Gly Phe Arg Gly Pro Ser Cys
            340                 345                 350
Gln Arg Pro Cys Pro Pro Gly Arg Tyr Gly Lys Arg Cys Val Pro Cys
        355                 360                 365
Lys Cys Asn Asn His Ser Ser Cys His Pro Ser Asp Gly Thr Cys Ser
    370                 375                 380
Cys Leu Ala Gly Trp Thr Gly Pro Asp Cys Ser Glu Ser Cys Pro Pro
385                 390                 395                 400
Gly His Trp Gly Leu Lys Cys Ser Gln Pro Cys Gln Cys His His Gly
                405                 410                 415
Ala Thr Cys His Pro Gln Asp Gly Ser Cys Val Cys Ile Pro Gly Trp
            420                 425                 430
Thr Gly Pro Asn Cys Ser Glu Gly Cys Pro Ser Arg Met Phe Gly Val
        435                 440                 445
Asn Cys Ser Gln Leu Cys Gln Cys Asp Pro Gly Glu Met Cys His Pro
    450                 455                 460
Glu Thr Gly Ala Cys Val Cys Pro Pro Gly His Ser Gly Ala His Cys
465                 470                 475                 480
Lys Val Gly Ser Gln Glu Ser Phe Thr Ile Met Pro Thr Ser Pro Val
                485                 490                 495
Ile His Asn Ser
            500

      <210> 123
      <211> 24
      <212> PRT
      <213> Rattus sp.

      <400> 123
Leu Gly Ala Val Ile Gly Ile Ala Val Leu Gly Thr Leu Val Val Ala
 1               5                  10                  15
Leu Val Ala Leu Phe Ile Gly Tyr
            20

      <210> 124
      <211> 112
      <212> PRT
      <213> Rattus sp.

      <400> 124
Arg His Trp Gln Lys Gly Lys Glu His Glu His Leu Ala Val Ala Tyr
 1               5                  10                  15
Ser Thr Gly Arg Leu Asp Gly Ser Asp Tyr Val Met Pro Asp Val Ser
            20                  25                  30
Pro Ser Tyr Ser His Tyr Tyr Ser Asn Pro Ser Tyr His Thr Leu Ser
        35                  40                  45
Gln Cys Ser Pro Asn Pro Pro Pro Pro Asn Lys Ile Pro Gly Ser Gln
    50                  55                  60
Leu Phe Val Ser Ser Gln Ala Ser Glu Arg Pro Asn Arg Asn His Gly
65                  70                  75                  80
Arg Asp Asn His Ala Thr Leu Pro Ala Asp Trp Lys His Arg Arg Glu
                85                  90                  95
Ser His Asp Arg Ala Phe Leu Arg His Gln Pro Pro Gly Pro Lys Val
```

100 105 110

<210> 125
<211> 28
<212> PRT
<213> Homo sapiens

<400> 125
Met Ala Pro Ala Arg Ala Gly Phe Cys Pro Leu Leu Leu Leu Leu Leu
1 5 10 15
Leu Gly Leu Trp Val Ala Glu Ile Pro Val Ser Ala
20 25

<210> 126
<211> 128
<212> PRT
<213> Homo sapiens

<400> 126
Lys Pro Lys Gly Met Thr Ser Ser Gln Trp Phe Lys Ile Gln His Met
1 5 10 15
Gln Pro Ser Pro Gln Ala Cys Asn Ser Ala Met Lys Asn Ile Asn Lys
20 25 30
His Thr Lys Arg Cys Lys Asp Leu Asn Thr Phe Leu His Glu Pro Phe
35 40 45
Ser Ser Val Ala Ala Thr Cys Gln Thr Pro Lys Ile Ala Cys Lys Asn
50 55 60
Gly Asp Lys Asn Cys His Gln Ser His Gly Pro Val Ser Leu Thr Met
65 70 75 80
Cys Lys Leu Thr Ser Gly Lys Tyr Pro Asn Cys Arg Tyr Lys Glu Lys
85 90 95
Arg Gln Asn Lys Ser Tyr Val Val Ala Cys Lys Pro Pro Gln Lys Lys
100 105 110
Asp Ser Gln Gln Phe His Leu Val Pro Val His Leu Asp Arg Val Leu
115 120 125

<210> 127
<211> 19
<212> PRT
<213> Homo sapiens

<400> 127
Met Pro Leu Leu Thr Leu Tyr Leu Leu Leu Phe Trp Leu Ser Gly Tyr
1 5 10 15
Ser Ile Ala

<210> 128
<211> 286
<212> PRT
<213> Homo sapiens

<400> 128
Thr Gln Ile Thr Gly Pro Thr Thr Val Asn Gly Leu Glu Arg Gly Ser
1 5 10 15
Leu Thr Val Gln Cys Val Tyr Arg Ser Gly Trp Glu Thr Tyr Leu Lys
20 25 30
Trp Trp Cys Arg Gly Ala Ile Trp Arg Asp Cys Lys Ile Leu Val Lys

35 40 45
Thr Ser Gly Ser Glu Gln Glu Val Lys Arg Asp Arg Val Ser Ile Lys
50 55 60
Asp Asn Gln Lys Asn Arg Thr Phe Thr Val Thr Met Glu Asp Leu Met
65 70 75 80
Lys Thr Asp Ala Asp Thr Tyr Trp Cys Gly Ile Glu Lys Thr Gly Asn
85 90 95
Asp Leu Gly Val Thr Val Gln Val Thr Ile Asp Pro Ala Ser Thr Pro
100 105 110
Ala Pro Thr Thr Pro Thr Ser Thr Thr Phe Thr Ala Pro Val Thr Gln
115 120 125
Glu Glu Thr Ser Ser Ser Pro Thr Leu Thr Gly His His Leu Asp Asn
130 135 140
Arg His Lys Leu Leu Lys Leu Ser Val Leu Leu Pro Leu Ile Phe Thr
145 150 155 160
Ile Leu Leu Leu Leu Leu Val Ala Ala Ser Leu Leu Ala Trp Arg Met
165 170 175
Met Lys Tyr Gln Gln Lys Ala Ala Gly Met Ser Pro Glu Gln Val Leu
180 185 190
Gln Pro Leu Glu Gly Asp Leu Cys Tyr Ala Asp Leu Thr Leu Gln Leu
195 200 205
Ala Gly Thr Ser Pro Arg Lys Ala Thr Thr Lys Leu Ser Ser Ala Gln
210 215 220
Val Asp Gln Val Glu Val Glu Tyr Val Thr Met Ala Ser Leu Pro Lys
225 230 235 240
Glu Asp Ile Ser Tyr Ala Ser Leu Thr Leu Gly Ala Glu Asp Gln Glu
245 250 255
Pro Thr Tyr Cys Asn Met Gly His Leu Ser Ser His Leu Pro Gly Arg
260 265 270
Gly Pro Glu Glu Pro Thr Glu Tyr Ser Thr Ile Ser Arg Pro
275 280 285

<210> 129
<211> 150
<212> PRT
<213> Homo sapiens

<400> 129
Thr Gln Ile Thr Gly Pro Thr Thr Val Asn Gly Leu Glu Arg Gly Ser
1 5 10 15
Leu Thr Val Gln Cys Val Tyr Arg Ser Gly Trp Glu Thr Tyr Leu Lys
20 25 30
Trp Trp Cys Arg Gly Ala Ile Trp Arg Asp Cys Lys Ile Leu Val Lys
35 40 45
Thr Ser Gly Ser Glu Gln Glu Val Lys Arg Asp Arg Val Ser Ile Lys
50 55 60
Asp Asn Gln Lys Asn Arg Thr Phe Thr Val Thr Met Glu Asp Leu Met
65 70 75 80
Lys Thr Asp Ala Asp Thr Tyr Trp Cys Gly Ile Glu Lys Thr Gly Asn
85 90 95
Asp Leu Gly Val Thr Val Gln Val Thr Ile Asp Pro Ala Ser Thr Pro
100 105 110
Ala Pro Thr Thr Pro Thr Ser Thr Thr Phe Thr Ala Pro Val Thr Gln
115 120 125
Glu Glu Thr Ser Ser Ser Pro Thr Leu Thr Gly His His Leu Asp Asn
130 135 140
Arg His Lys Leu Leu Lys
145 150

<210> 130
<211> 24
<212> PRT
<213> Homo sapiens

<400> 130
Leu Ser Val Leu Leu Pro Leu Ile Phe Thr Ile Leu Leu Leu Leu Leu
1 5 10 15
Val Ala Ala Ser Leu Leu Ala Trp
20

<210> 131
<211> 112
<212> PRT
<213> Homo sapiens

<400> 131
Arg Met Met Lys Tyr Gln Gln Lys Ala Ala Gly Met Ser Pro Glu Gln
1 5 10 15
Val Leu Gln Pro Leu Glu Gly Asp Leu Cys Tyr Ala Asp Leu Thr Leu
20 25 30
Gln Leu Ala Gly Thr Ser Pro Arg Lys Ala Thr Thr Lys Leu Ser Ser
35 40 45
Ala Gln Val Asp Gln Val Glu Val Glu Tyr Val Thr Met Ala Ser Leu
50 55 60
Pro Lys Glu Asp Ile Ser Tyr Ala Ser Leu Thr Leu Gly Ala Glu Asp
65 70 75 80
Gln Glu Pro Thr Tyr Cys Asn Met Gly His Leu Ser Ser His Leu Pro
85 90 95
Gly Arg Gly Pro Glu Glu Pro Thr Glu Tyr Ser Thr Ile Ser Arg Pro
100 105 110

<210> 132
<211> 21
<212> PRT
<213> Homo sapiens

<400> 132
Met Asp His Cys Gly Ala Leu Phe Leu Cys Leu Cys Leu Leu Thr Leu
1 5 10 15
Gln Asn Ala Thr Thr
20

<210> 133
<211> 507
<212> PRT
<213> Homo sapiens

<400> 133
Glu Thr Trp Glu Glu Leu Leu Ser Tyr Met Glu Asn Met Gln Val Ser
1 5 10 15
Arg Gly Arg Ser Ser Val Phe Ser Ser Arg Gln Leu His Gln Leu Glu
20 25 30
Gln Met Leu Leu Asn Thr Ser Phe Pro Gly Tyr Asn Leu Thr Leu Gln
35 40 45
Thr Pro Thr Ile Gln Ser Leu Ala Phe Lys Leu Ser Cys Asp Phe Ser
50 55 60
Gly Leu Ser Leu Thr Ser Ala Thr Leu Lys Arg Val Pro Gln Ala Gly

```
65                  70                  75                  80
Gly Gln His Ala Arg Gly Gln His Ala Met Gln Phe Pro Ala Glu Leu
                85                  90                  95
Thr Arg Asp Ala Cys Lys Thr Arg Pro Arg Glu Leu Arg Leu Ile Cys
            100                 105                 110
Ile Tyr Phe Ser Asn Thr His Phe Phe Lys Asp Glu Asn Asn Ser Ser
        115                 120                 125
Leu Leu Asn Asn Tyr Val Leu Gly Ala Gln Leu Ser His Gly His Val
    130                 135                 140
Asn Asn Leu Arg Asp Pro Val Asn Ile Ser Phe Trp His Asn Gln Ser
145                 150                 155                 160
Leu Glu Gly Tyr Thr Leu Thr Cys Val Phe Trp Lys Glu Gly Ala Arg
                165                 170                 175
Lys Gln Pro Trp Gly Gly Trp Ser Pro Glu Gly Cys Arg Thr Glu Gln
            180                 185                 190
Pro Ser His Ser Gln Val Leu Cys Arg Cys Asn His Leu Thr Tyr Phe
        195                 200                 205
Ala Val Leu Met Gln Leu Ser Pro Ala Leu Val Pro Ala Glu Leu Leu
    210                 215                 220
Ala Pro Leu Thr Tyr Ile Ser Leu Val Gly Cys Ser Ile Ser Ile Val
225                 230                 235                 240
Ala Ser Leu Ile Thr Val Leu Leu His Phe His Phe Arg Lys Gln Ser
                245                 250                 255
Asp Ser Leu Thr Arg Ile His Met Asn Leu His Ala Ser Val Leu Leu
            260                 265                 270
Leu Asn Ile Ala Phe Leu Leu Ser Pro Ala Phe Ala Met Ser Pro Val
        275                 280                 285
Pro Gly Ser Ala Cys Thr Ala Leu Ala Ala Ala Leu His Tyr Ala Leu
    290                 295                 300
Leu Ser Cys Leu Thr Trp Met Ala Ile Glu Gly Phe Asn Leu Tyr Leu
305                 310                 315                 320
Leu Leu Gly Arg Val Tyr Asn Ile Tyr Ile Arg Arg Tyr Val Phe Lys
                325                 330                 335
Leu Gly Val Leu Gly Trp Gly Ala Pro Ala Leu Leu Val Leu Leu Ser
            340                 345                 350
Leu Ser Val Lys Ser Ser Val Tyr Gly Pro Cys Thr Ile Pro Val Phe
        355                 360                 365
Asp Ser Trp Glu Asn Gly Thr Gly Phe Gln Asn Met Ser Ile Cys Trp
    370                 375                 380
Val Arg Ser Pro Val Val His Ser Val Leu Val Met Gly Tyr Gly Gly
385                 390                 395                 400
Leu Thr Ser Leu Phe Asn Leu Val Val Leu Ala Trp Ala Leu Trp Thr
                405                 410                 415
Leu Arg Arg Leu Arg Glu Arg Ala Asp Ala Pro Ser Val Arg Ala Cys
            420                 425                 430
His Asp Thr Val Thr Val Leu Gly Leu Thr Val Leu Leu Gly Thr Thr
        435                 440                 445
Trp Ala Leu Ala Phe Phe Ser Phe Gly Val Phe Leu Leu Pro Gln Leu
    450                 455                 460
Phe Leu Phe Thr Ile Leu Asn Ser Leu Tyr Gly Phe Phe Leu Phe Leu
465                 470                 475                 480
Trp Phe Cys Ser Gln Arg Cys Arg Ser Glu Ala Glu Ala Lys Ala Gln
                485                 490                 495
Ile Glu Ala Phe Ser Ser Ser Gln Thr Thr Gln
            500                 505
```

<210> 134
<211> 223

<212> PRT
<213> Homo sapiens

<400> 134
Glu Thr Trp Glu Glu Leu Leu Ser Tyr Met Glu Asn Met Gln Val Ser
1 5 10 15
Arg Gly Arg Ser Ser Val Phe Ser Ser Arg Gln Leu His Gln Leu Glu
20 25 30
Gln Met Leu Leu Asn Thr Ser Phe Pro Gly Tyr Asn Leu Thr Leu Gln
35 40 45
Thr Pro Thr Ile Gln Ser Leu Ala Phe Lys Leu Ser Cys Asp Phe Ser
50 55 60
Gly Leu Ser Leu Thr Ser Ala Thr Leu Lys Arg Val Pro Gln Ala Gly
65 70 75 80
Gly Gln His Ala Arg Gly Gln His Ala Met Gln Phe Pro Ala Glu Leu
85 90 95
Thr Arg Asp Ala Cys Lys Thr Arg Pro Arg Glu Leu Arg Leu Ile Cys
100 105 110
Ile Tyr Phe Ser Asn Thr His Phe Phe Lys Asp Glu Asn Asn Ser Ser
115 120 125
Leu Leu Asn Asn Tyr Val Leu Gly Ala Gln Leu Ser His Gly His Val
130 135 140
Asn Asn Leu Arg Asp Pro Val Asn Ile Ser Phe Trp His Asn Gln Ser
145 150 155 160
Leu Glu Gly Tyr Thr Leu Thr Cys Val Phe Trp Lys Glu Gly Ala Arg
165 170 175
Lys Gln Pro Trp Gly Gly Trp Ser Pro Glu Gly Cys Arg Thr Glu Gln
180 185 190
Pro Ser His Ser Gln Val Leu Cys Arg Cys Asn His Leu Thr Tyr Phe
195 200 205
Ala Val Leu Met Gln Leu Ser Pro Ala Leu Val Pro Ala Glu Leu
210 215 220

<210> 135
<211> 25
<212> PRT
<213> Homo sapiens

<400> 135
Leu Ala Pro Leu Thr Tyr Ile Ser Leu Val Gly Cys Ser Ile Ser Ile
1 5 10 15
Val Ala Ser Leu Ile Thr Val Leu Leu
20 25

<210> 136
<211> 20
<212> PRT
<213> Homo sapiens

<400> 136
Leu His Ala Ser Val Leu Leu Leu Asn Ile Ala Phe Leu Leu Ser Pro
1 5 10 15
Ala Phe Ala Met
20

<210> 137
<211> 21
<212> PRT

<213> Homo sapiens

<400> 137
Tyr Ala Leu Leu Ser Cys Leu Thr Trp Met Ala Ile Glu Gly Phe Asn
1 5 10 15
Leu Tyr Leu Leu Leu
20

<210> 138
<211> 19
<212> PRT
<213> Homo sapiens

<400> 138
Leu Gly Val Leu Gly Trp Gly Ala Pro Ala Leu Leu Val Leu Leu Ser
1 5 10 15
Leu Ser Val

<210> 139
<211> 25
<212> PRT
<213> Homo sapiens

<400> 139
Val Leu Val Met Gly Tyr Gly Gly Leu Thr Ser Leu Phe Asn Leu Val
1 5 10 15
Val Leu Ala Trp Ala Leu Trp Thr Leu
20 25

<210> 140
<211> 21
<212> PRT
<213> Homo sapiens

<400> 140
Val Thr Val Leu Gly Leu Thr Val Leu Leu Gly Thr Thr Trp Ala Leu
1 5 10 15
Ala Phe Phe Ser Phe
20

<210> 141
<211> 20
<212> PRT
<213> Homo sapiens

<400> 141
Leu Phe Leu Phe Thr Ile Leu Asn Ser Leu Tyr Gly Phe Phe Leu Phe
1 5 10 15
Leu Trp Phe Cys
20

<210> 142
<211> 24
<212> PRT
<213> Homo sapiens

<400> 142

Ser Gln Arg Cys Arg Ser Glu Ala Glu Ala Lys Ala Gln Ile Glu Ala
1 5 10 15
Phe Ser Ser Ser Gln Thr Thr Gln
20

<210> 143
<211> 16
<212> PRT
<213> Homo sapiens

<400> 143
Ser Pro Val Pro Gly Ser Ala Cys Thr Ala Leu Ala Ala Ala Leu His
1 5 10 15

<210> 144
<211> 37
<212> PRT
<213> Homo sapiens

<400> 144
Lys Ser Ser Val Tyr Gly Pro Cys Thr Ile Pro Val Phe Asp Ser Trp
1 5 10 15
Glu Asn Gly Thr Gly Phe Gln Asn Met Ser Ile Cys Trp Val Arg Ser
20 25 30
Pro Val Val His Ser
35

<210> 145
<211> 7
<212> PRT
<213> Homo sapiens

<400> 145
Gly Val Phe Leu Leu Pro Gln
1 5

<210> 146
<211> 17
<212> PRT
<213> Homo sapiens

<400> 146
His Phe His Phe Arg Lys Gln Ser Asp Ser Leu Thr Arg Ile His Met
1 5 10 15
Asn

<210> 147
<211> 14
<212> PRT
<213> Homo sapiens

<400> 147
Gly Arg Val Tyr Asn Ile Tyr Ile Arg Arg Tyr Val Phe Lys
1 5 10

<210> 148
<211> 18

<212> PRT
<213> Homo sapiens

<400> 148
Arg Arg Leu Arg Glu Arg Ala Asp Ala Pro Ser Val Arg Ala Cys His
 1   5   10   15
Asp Thr

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/18198

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C07K 14/47; C07H 21/04; C12N 15/63, 1/2; C12P 21/02
US CL : 530/350; 536/23.5; 435/320.1, 252.3, 361, 69.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/350; 536/23.5; 435/320.1, 252.3, 361, 69.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Commercial Sequence Databases: GenEmbl, EST, Issued_Patents_NA, N_Geneseq_36, PIR_64, SwissProt_38, A_Geneseq_36,Issued_Patents_AA, SPTREMBL_12

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X — Y	Database EST, AN AQ588144, ZHOU et al. 'CITBI-E1-2644L24.TF CITBI-E1 Homo sapiens genomic clone 2644L24, genomic survey sequence'. 07 June 1999, see attached alignment showing 100% identical match to nucleotides 88-481 of SEQ ID NO: 1 (394 nucleotides total).	1, 3, 5 — 2, 4, 6-10 and 12
A	Database SPTREMBL_12, AN Q28396, RICHARDSON et al. 'Type II Collagen from Equus caballus (Horse)'. 01 November 1996. Polypeptide 25.7% identical to the amino acid sequence of SEQ ID NO:2, see attached alignment, Nov. 1, 1996.	1-10 and 12

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 SEPTEMBER 2000	02 OCT 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer EILEEN B. O'HARA Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/18198

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-10 and 12

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s)1-10 and 12, in so far as they are drawn to Intercept 340, polynucleotides of SEQ ID NOS: 1 and 3, vector, host cell, method of producing a protein recombinantly and protein of SEQ ID NO: 2.
Groups II-VII, claim(s) 1-10 and 12, in so far as they are drawn to the next six polynucleotides of distinct cDNA clones and encoded proteins, identified as Mango 003, Mango 347, Tango 272, Tango 295, Tango 354 and Tango 378, as listed in Tables 1 and 2.
Groups VIII-XIV, claim(s) 11 and 15, in so far as they are drawn to antibodies to one of the seven proteins listed above.
Groups XV-XXI, claims 13, 14, 19, 20 and 22, in so far as they are drawn to a method for detecting the presence of in a sample or identifying a compound which binds to or modulates the activity of a polypeptide of one of the seven proteins listed above.
Groups XXII-XXVII, claims 16 and 17, in so far as they are drawn to a method for detecting the nucleic acids of one of the seven cDNA clones listed above.
Groups XXIX-XXXV, claim 18, in so far as it is drawn to a kit comprising a compound of unspecified constitution which selectively binds to a nucleic acid molecule of the seven cDNA clones listed above.
Groups XXXVI-XLII, claim 21, in so far as it is drawn to a method for modulating the activity of one of the seven proteins listed above.

The inventions listed as Groups I-XLII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Group I corresponds to the first invention wherein the first product is the polynucleotide and the first method of using is the method of making the protein. Note that there is no method of making the polynucleotide. The invention also includes the protein made. Each of groups II-VII does not share the same or corresponding special technical feature because each group is drawn to a different polynucleotide and encoded protein, and each of groups VIII-XLII does not share the same or corresponding special technical feature because each group is drawn to different compounds or methods of using the seven polynucleotides and encoded proteins. This Authority therefore considers that the several inventions do not share a special technical feature within the meaning of PCT Rule 13.2 and thus do not relate to a single general inventive concept within the meaning of PCT Rule 13.1.